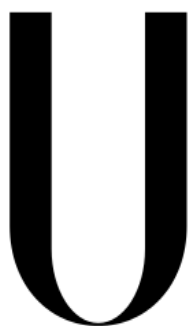


**UNIVERSIDADE DE LISBOA**

**FACULDADE DE CIÊNCIAS**

**DEPARTAMENTO DE ENGENHARIA GEOGRÁFICA**

**GEOFÍSICA E ENERGIA**



**LISBOA**

---

UNIVERSIDADE  
DE LISBOA

**GENERALIZAÇÃO CARTOGRÁFICA COM RECURSO A  
INTELIGÊNCIA ARTIFICIAL**

**José António Travanca Lopes**

Tese orientada pelo Professor Doutor João Catalão Fernandes

**DOUTORAMENTO EM CIÊNCIAS GEOFÍSICAS E DA  
GEOINFORMAÇÃO**

**(Sistemas de Informação Geográfica)**

**2015**

DOCUMENTO ESPECIALMENTE ELABORADO PARA A OBTENÇÃO DO GRAU DE DOUTOR









# Resumo

As agências cartográficas nacionais estão institucionalmente comprometidas com a produção multi-escala que exige a manutenção de enormes volumes de dados de diferentes resoluções espaciais e temporais de difícil gestão e manutenção. A alternativa a este ciclo crescente do volume de dados reside na implementação de sistemas de generalização automática que permitam reduzir a produção e manutenção de uma única série cartográfica de maior exatidão geométrica e semântica.

Neste trabalho apresenta-se o processo de desenvolvimento e implementação de uma infraestrutura de generalização cartográfica automática de linhas baseada numa estratégia de estrangimento e refinamento iterativo com recurso a algoritmos de inteligência artificial. Considera-se um algoritmo paramétrico de simplificação e suavização de linhas aplicado ao caso particular de curvas de nível. A automatização do processo de generalização cartográfica de linhas passa por seleccionar de forma automática o parâmetro a utilizar no algoritmo, a tensão a aplicar à linha. A proposta consiste no recurso a técnicas de inteligência artificial para seleção do valor ótimo do parâmetro tensão. É utilizada uma Rede Neuronal, uma Árvore de Decisão e uma Árvore de Classificação e Regressão, para o cálculo do valor da tensão, obtidos os valores preditos por estes métodos é escolhido o ótimo entre eles através de um Agente que executa um esquema de leilão. A metodologia aplicada inicia-se com a caracterização numérica das curvas de nível, calculando entre outras, a dimensão fractal, comprimento, área, angularidade, a partir das quais é calculado o valor da tensão a usar no algoritmo, para cada curva de nível. Por fim são apresentados dois algoritmos para contextualização da generalização cartográfica de curvas de nível com os pontos de cota, vértices geodésicos e linhas de água, respeitando as leis de Brisson.

**Palavras-chave:** Generalização cartográfica, Algoritmo, Inteligência Artificial, Curva de nível, Contextualização.





# Abstract

The National Mapping Agencies are institutionally compromised with the multi-scale production that requires the maintenance of enormous amount of data with different space and time resolutions increasing the management and maintenance. The alternative to this cycle of data increasing must be the implementation of automatic cartographic generalization systems that will lead to a significant reduction in the data amount and data maintenance in a unique cartographic series of higher geometric and semantic accuracy.

In this work, a new cartographic generalization infrastructure is presented. The infrastructure was developed for the cartographic generalization of lines and is based on the constraint and iterative strategy using artificial intelligence algorithms. We have considered a parametric algorithm for line simplification and smooth applied to the particular case of contour lines. The automation of the process is based on the intelligent selection of the parameter to be used for the line generalization. The parameter is the tension to be applied to the line that is elected among all possible values using artificial intelligence techniques. A Neuronal Net, a Decision Tree and a Classification and Regression Tree, are used for the selection of the tension to be applied to the curve that are further auctioned through an Agent that selects the best tension value. The applied methodology is initiated with the numerical characterization of the contour lines, calculating among others, the fractal dimension, length, area, angularity, from which the value of the tension is calculated to use in the algorithm, for each contour line. Finally, two algorithms for the contextualization of the generalized contour lines with the altimetry points are presented, geodesic vertices and water lines.

**Key Words:** Cartographic generalization, Algorithm, Artificial Intelligence, Line Contour, Contextualization.





# Agradecimentos

A investigação do tema da generalização cartográfica é um desafio grandioso. Uma tarefa destas dimensões, nos dias de hoje, com toda a tecnologia disponível e o conhecimento que envolve, deve ser abordada por equipas multidisciplinares pela sua complexidade e abrangência.

Quando se abraça um desafio destes, é necessário que se interrogue até onde poderemos ir, o céu é realmente o limite?

Esta investigação e correspondente dissertação, não teria sido possível sem o apoio de diversas pessoas e organizações. Estou profundamente agradecido a todos e gostaria de agradecer a todos em particular. Primeiro, agradeço ao Exército Português e à Academia Militar pelo reconhecimento e oportunidade de efetuar esta investigação. Em segundo lugar ao Instituto Geográfico do Exército por acolher e apoiar esta iniciativa e facultar toda a informação digital, recursos humanos e tecnológicos, para realização deste projeto.

Expresso os meus sinceros agradecimentos à Sra Dra Anne Ruas pelo apoio incondicional prestado durante a visita ao IGN Francês, a toda a ajuda na elaboração dos artigos publicados e na clarificação de ideias. Ao Dr. Gaffuri pela disponibilidade, debate de ideias e apoio durante a visita ao IGN Francês.

Ao professor Catalão orientador desta dissertação, agradeço a grande disponibilidade, o incentivo e a persistência, a colaboração e a amizade sempre manifestadas, pela orientação e apoio na elaboração da tese, dos artigos e rumo dos trabalhos.

Ao Cor Estrela Soares, diretor do IGeoE, pelo permanente apoio em informação cartográfica e disponibilização de meios humanos para a realização de todo o trabalho. Aos operadores do IGeoE, nomeadamente ao Capitão Costa, SChf Antunes, SAjd Alves, SAjd Sendim, 1ºSarg Castanheira e Tec Inf Anabela pelas revisões das folhas generalizadas automaticamente, pelos conselhos e propostas de melhorias ao algoritmo e metodologia.

Aos Tenente Coronel Galindro, Tenente Coronel Valentim, Tenente Coronel Rocha, Tenente Coronel Caravela e Major Valente pelas revisões e propostas de melhoria a esta tese.

Á Eng. Sónia e D. Dulce pelo apoio prestado e pelo encorajamento. Um profundo e sincero agradecimento à minha família, esposa e filha, que privei de muitos dias da minha companhia mesmo em dias chamados de férias, feriados ou fim de semana, dos quais necessitava do tempo para trabalhar. Ao apoio e estabilidade familiar imprescindível para levar a bom termo uma tarefa hercúlea destas dimensões.

A todos os que contribuíram direta ou indiretamente para a concretização deste trabalho, muito obrigado.

# ÍNDICE

Resumo .....	i
Abstract .....	iii
Agradecimentos .....	v
Índice .....	vii
Lista de Tabelas .....	xi
Lista de Figuras .....	xiii
Lista de Acrónimos .....	xvii

## Capítulo 1 – Introdução

1.1 - Introdução .....	1
1.2 - Objetivos .....	3
1.3 - Organização da dissertação .....	3

## Capítulo 2 – Generalização cartográfica

2.1 - Generalização cartográfica .....	5
2.2 - Da generalização cartográfica manual à digital .....	6
2.3 - Modelos conceptuais de generalização .....	9
2.3.1 - Modelo de Ratajski .....	9
2.3.2 - Modelo de Brassel e Weibel .....	10
2.3.3 - Modelo de McMaster e Shea .....	12
2.3.4 - Modelo de Grünreich .....	14
2.3.5 - Modelo AGENT .....	16
2.4 - Operadores e algoritmos de generalização .....	18
2.4.1 - Classificação .....	19
2.4.2 - Simbolização .....	20
2.4.3 - Agregação .....	21
2.4.4 - Seleção .....	21
2.4.5 - Simplificação .....	22
2.4.6 - Suavização .....	23
2.4.7 - Realce .....	23
2.4.8 - Fractalização .....	24
2.4.9 - Fusão .....	24
2.4.10 - Amalgamação .....	25
2.4.11 - Colapso .....	26
2.4.12 - Exagero .....	26
2.4.13 - Tipificação .....	27
2.4.14 - Deslocamento .....	28
2.4.15 - Refinamento .....	28
2.4.16 - Anamorfose .....	29
2.5 - Generalização de linhas .....	30
2.5.1 - Segmentação de linhas .....	32
2.5.2 - Caracterização de linhas .....	34
2.5.3 - Classificação de linhas .....	35
2.6 - Estratégias para a generalização automática .....	36

## Capítulo 3 – Conhecimento e IA

3.1 - Introdução .....	41
3.2 - Conhecimento em generalização cartográfica .....	41
3.2.1 - Técnicas de aquisição de conhecimento .....	43
a) <i>Perícia humana e análise de documentos</i> .....	43
b) <i>Aprendizagem máquina</i> .....	44
b1) <i>Aprendizagem Supervisionada</i> .....	44
b2) <i>Aprendizagem não Supervisionada</i> .....	45
3.3 - Representação do conhecimento .....	45
3.4 - Inteligência Artificial .....	48
3.5 - Redes Neurais .....	48
3.5.1 - Topologia de Redes Neurais .....	50
3.5.2 - Aprendizagem em Redes Neurais .....	51
3.5.3 - Tamanho da Rede Neuronal Artificial .....	54
3.6 - Árvores de decisão .....	55
3.6.1 - Representação de uma árvore de decisão .....	56
3.6.2 - Problemas apropriados para a aprendizagem em árvores de decisão .....	58
3.6.3 - Regras de decisão .....	59
3.7 - Árvores de classificação e regressão .....	59
3.8 - Considerações sobre os algoritmos C4.5 e CART .....	61
3.9 - Sobre ajustamento dos dados .....	61
3.10 - O Agente .....	63

## Capítulo 4 – Generalização Cartográfica de linhas

4.1 - Introdução .....	67
4.2 - Caracterização das séries cartográficas envolvidas .....	68
4.2.1 - A Base de Dados Geográficos de média escala .....	68
4.2.2 - Série M888 à escala 25k .....	69
4.2.3 - Série M782 à escala 50k .....	70
4.3 - Generalização modelo .....	70
4.3.1 - Seleção das curvas de nível .....	70
4.3.2 - Seleção dos pontos de cota .....	71
4.3.3 - Classificação e seleção das linhas de água .....	72
4.3.4 - Classificação das linhas de água pelo método de Strahler .....	74
4.3.5 - Classificação das linhas de água pelo método de Shreve .....	74
4.3.6 - Classificação e seleção pelo método do IGeoE .....	75
4.4 - Algoritmos de simplificação de linhas .....	78
4.4.1 - Algoritmos com parâmetros geométricos como critério .....	79
4.4.2 - Algoritmos com funções de parâmetros geométricos como critério .....	82
4.4.3 - Algoritmos de suavização de linhas .....	84
4.5 - Algoritmo TLopes .....	85
4.6 - Comparação com outros algoritmos de generalização de linhas .....	87

## Capítulo 5 - Generalização cartográfica de linhas com recurso a IA

5.1 - Introdução .....	95
5.2 - Estratégia de processamento .....	96
5.3 - Cálculo dos atributos das curvas de nível .....	97
5.4 - Uso da IA na generalização de curvas de nível .....	100
5.5 - Treino dos classificadores .....	103



<b>5.6 - Validação .....</b>	<b>109</b>
5.6.1 - Validação qualitativa .....	111
5.6.2 - Testes de sensibilidade do modelo .....	114
5.6.3 - Validação quantitativa .....	117
<b>5.7 - Contextualização em Generalização .....</b>	<b>124</b>
5.7.1 - Generalização contextual entre curvas de nível e pontos de cota .....	124
5.7.2 - Generalização contextual entre curvas de nível e linhas de água .....	125
<b>5.8 - Comparação com outros métodos .....</b>	<b>128</b>

## **Capítulo 6 – Conclusões**

<b>6.1 - Conclusão .....</b>	<b>131</b>
------------------------------	------------

<b>Referências Bibliográficas .....</b>	<b>135</b>
---	------------



# Lista de Tabelas

Tabela 5.1 – Parâmetros característicos e sua descrição .....	100
Tabela 5.2 – Estatística da caracterização de linhas .....	104
Tabela 5.3 – Importância das variáveis no processo de classificação .....	105
Tabela 5.4 – Matriz de confusão para a Rede Neuronal .....	109
Tabela 5.5 – Matriz de confusão para a Árvore C5 .....	109
Tabela 5.6 – Matriz de confusão para a Árvore C&R .....	109
Tabela 5.7 – Matriz de confusão depois do Agente .....	109
Tabela 5.8 – Matriz de confusão para a Rede Neuronal .....	117
Tabela 5.9 – Matriz de confusão para a Árvore C5 .....	117
Tabela 5.10 – Matriz de confusão para a Árvore C&R .....	117
Tabela 5.11 – Matriz de confusão depois do Agente .....	117
Tabela 5.12 – Comparação entre as curvas de nível originais, escala 25k, e generalizadas manualmente para a escala 50k, folha 279 .....	118
Tabela 5.12a – Diferenças entre as curvas de nível originais, escala 25k, e generalizadas manualmente para a escala 50k, folha 279 .....	118
Tabela 5.13 – Comparação entre as curvas de nível na escala 50k, generalizadas manualmente e automaticamente 50k, folha 279 .....	119
Tabela 5.13a – Diferenças entre as curvas de nível na escala 50k, generalizadas manualmente e automaticamente, folha 279 .....	120
Tabela 5.14 – Comparação entre a dimensão fractal das curvas de nível da folha 25k e 50k, folha 279 .	121
Tabela 5.15 – Comparação entre os comprimentos das curvas de nível da folha 25k e 50k generaliza- das manualmente e automaticamente, folha 279 .....	123
Tabela 5.16 – Matriz de confusão das restantes folhas utilizadas .....	123



# Lista de Figuras

Figura 2.1 – Modelo da generalização proposto por Ratajski (adaptado de Cecconi 2003) .....	10
Figura 2.2 – Modelo conceptual de Brassel e Weibel, 1988, para a generalização, adaptada de (Mackaness et al. 2007) .....	11
Figura 2.3 – Modelo conceptual de McMaster e Shea (1992) para a generalização em ambiente digital, adaptada de (Mackaness et al. 2007) .....	13
Figura 2.4 – Modelo de generalização de Grünreich (1992), adaptado de Foerster et al. (2010) .....	15
Figura 2.5 – Modelo de generalização AGENT, relação entre agentes micro, meso e macro .....	17
Figura 2.6 – Modelo de generalização AGENT, proposto por (Galanda e Weibel 2002) .....	18
Figura 2.7 – O operador classificação .....	20
Figura 2.8 – O operador simbolização .....	20
Figura 2.9 – O operador agregação .....	21
Figura 2.10 – O operador seleção .....	22
Figura 2.11 – O operador de simplificação .....	23
Figura 2.12 – O operador de suavização .....	23
Figura 2.13 – O operador de realce .....	24
Figura 2.14 – O operador de fractalização (adaptado de Falconer 1990) .....	24
Figura 2.15 – O operador de fusão .....	25
Figura 2.16 – O operador amalgamação .....	26
Figura 2.17 – O operador de colapso .....	26
Figura 2.18 – O operador de exagero .....	27
Figura 2.19 – O operador de tipificação .....	28
Figura 2.20 – O operador de deslocamento .....	28
Figura 2.21 – O operador de refinamento .....	29
Figura 2.22 – O operador anamorfose .....	29
Figura 2.23 – Cálculo das medidas de caracterização utilizada por Plazanet (1996) .....	34
Figura 2.24 – Plano geral da estratégia proposta por Ruas e Plazanet (adaptado de Ruas e Plazanet 1996) .....	38
Figura 3.1 – Módulos de um sistema baseado em conhecimento (adaptado de Kazemi et al. 2005) .....	47
Figura 3.2 – Esquema de unidade de processamento, McCulloch-Pitts .....	49
Figura 3.3 – Rede neuronal organizada em camadas. Neste caso tem uma camada de entrada e duas intermédias .....	51
Figura 3.4 – Algoritmo de treino, Regra Delta .....	53

Figura 3.5 – Exemplos de funções de ativação .....	53
Figura 3.6 – Representação de uma árvore de decisão e respectiva representação no espaço (adaptado de Gama et al. 2012) .....	57
Figura 4.1 – Temas ou Superclasses representadas na série M888, generalizadas para a série M782 ..	69
Figura 4.2 – Pontos de cota a apagar no interior do buffer, com centro no VG .....	72
Figura 4.3 – Classes pertencentes ao tema hidrografia .....	73
Figura 4.4 – Classificação de linhas de água pelo método de Strahler .....	74
Figura 4.5 – Classificação de linhas de água pelo método de Shreve .....	75
Figura 4.6 – Atribuição de pesos de forma ascendente .....	76
Figura 4.7 – Classificação e seleção das Linhas de água a representar na folha, escala 1/50 000 .....	77
Figura 4.8 – Seleção das linhas de água a representar na folha, escala 1/50 000 .....	78
Figura 4.9 – Seleção de pontos pelo algoritmo do N-ésimo ponto .....	79
Figura 4.10 – Seleção de pontos com a distância como critério .....	79
Figura 4.11 – Seleção de pontos com o ângulo como critério .....	80
Figura 4.12 – O princípio de funcionamento do algoritmo de Reumann-Witkam .....	80
Figura 4.13 – Algoritmo de Douglas-Peucker .....	81
Figura 4.14 – O princípio de funcionamento do algoritmo de Ansari-Delp .....	81
Figura 4.15 – O princípio de funcionamento do algoritmo de Visvalingham-Whyatt .....	82
Figura 4.16 – O princípio de funcionamento do algoritmo de Rosenfeld-Johnston .....	82
Figura 4.17 – O princípio de funcionamento do algoritmo baseado no rácio $d/S$ .....	82
Figura 4.18 – Algoritmo de Nakos-Mitropoulos, usando o rácio $L/S$ .....	83
Figura 4.19 – Algoritmo de Suavização .....	84
Figura 4.20 – Ação e reação da curva de nível ao sistema de forças aplicado .....	85
Figura 4.21a – Cálculo dos vectores $\vec{A}_{12}$ , $\vec{A}_{23}$ e pontos máximos e mínimos .....	86
Figura 4.21b – Cálculo dos vectores $\vec{B}_i$ perpendiculares ao vector $\vec{A}_{j,j+1}$ .....	86
Figura 4.21c – Exemplo de diferentes tensões aplicadas à curva de nível .....	86
Figura 4.22 – Interface do programa para a generalização das curvas de nível .....	87
Figura 4.23 – Comparação com o algoritmo de Lang .....	88
Figura 4.24 – Comparação com o algoritmo de Douglas-Peucker .....	88
Figura 4.25 – Comparação com o algoritmo de Douglas-Peucker. Preto – curva de nível ímpar original escala 25k, Vermelho - curva de nível par original escala 25k, Azul - curva de nível generalizada com algoritmo de Douglas-Peucker, Verde – curva de nível generalizada com algoritmo TLopes .....	89
Figura 4.26 – Comparação das linhas generalizadas manualmente com as generalizadas automaticamente com o algoritmo TLopes .....	90
Figura 4.27 – MDT pixel de 10 metros (classes altimetria) .....	90
Figura 4.28 – Extratos da folha com curvas de nível originais, escala 25k, de cor vermelha, versus curvas vectorizadas grid de 5 metros, cor verde, e vectorizadas após reamostragem para	

grid 10m cor azul.	
a) e c) incluindo as curvas originais;	
b) e d) apenas com as curvas generalizadas .....	91
<b>Figura 4.29</b> – Extratos da folha com curvas de nível originais, escala 25k, de cor azul, versus curvas vectorizadas, após reamostragem, de cor vermelha, generalizadas para a escala 50k .....	92
<b>Figura 4.30</b> – Comparação entre curvas generalizadas manualmente a vermelho; vectorizadas após reamostragem da grid para 10m a azul, e verde generalizadas com o algoritmo TLopes ...	92
<b>Figura 4.31</b> – Comparação entre as curvas de nível : Vermelho – original na escala 25k, Verde – Generalizadas semi-automaticamente com o algoritmo de Douglas Peucker e suavizadas com spline, Azul – Generalizadas com o algoritmo TLopes .....	93
<b>Figura 4.32</b> – Comparação das curvas generalizadas manualmente a vermelho e com o algoritmo TLopes a verde .....	94
<b>Figura 5.1</b> – Esquema geral da metodologia aplicada .....	96
<b>Figura 5.2</b> – Classes do tema altimetria .....	97
<b>Figura 5.3</b> – Passos para o cálculo da Dimensão Fractal de uma curva de nível .....	99
<b>Figura 5.4</b> – Fases do treino .....	101
<b>Figura 5.5</b> – Projeto Geral de atividades de Inteligência Artificial .....	102
<b>Figura 5.6</b> – Generalização das Curvas de Nível .....	103
<b>Figura 5.7</b> – Esquema geral das atividades de treino, validação e testes .....	104
<b>Figura 5.8</b> – Curvas de nível, da folha 309 à escala 25k .....	106
<b>Figura 5.9a</b> – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais .....	107
<b>Figura 5.9b</b> – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais .....	107
<b>Figura 5.9c</b> – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais .....	108
<b>Figura 5.9d</b> – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais .....	108
<b>Figura 5.10</b> – Resultados da metodologia aplicada a curvas de nível da folha 50 .....	110
<b>Figura 5.11</b> – Resultados da metodologia aplicada a curvas de nível, extratos das folhas utilizadas .....	111
<b>Figura 5.12</b> – Erros apontados pelos operadores da Secção de Controlo de Qualidade .....	113
<b>Figura 5.13</b> – Curvas de nível e linhas de água generalizadas para a escala 1/50 000 .....	115
<b>Figura 5.14</b> – Histogramas com os valores da tensão utilizada pelos operadores (Op1, Op2, Op3 e Op4) para uma mesma folha (N 279) .....	116
<b>Figura 5.15a</b> – Histogramas para os valores do $Comp/\sqrt{Area}$ para o conjunto de dados Originais, escala 25k .....	121
<b>Figura 5.15b</b> – Histogramas para os valores do $Comp/\sqrt{Area}$ para o conjunto de Dados generalizados manualmente .....	121
<b>Figura 5.15c</b> – Histogramas para os valores do $Comp/\sqrt{Area}$ para o conjunto de Dados	

generalizados automaticamente .....	121
<b>Figura 5.16</b> – Comparação entre os valores dos slivers, para o conjunto de dados originais e os correspondentes, generalizados manual e automaticamente. As curvas de nível foram agrupadas em 12 conjuntos, representados pelos raios no gráfico .....	122
<b>Figura 5.17</b> – Comparação entre os valores do comprimento, para o conjunto de dados generalizados manualmente e automaticamente .....	122
<b>Figura 5.18a</b> – Tensão aplicada á linha .....	125
<b>Figura 5.18b</b> – Query espacial, para detecção da violação da topologia .....	125
<b>Figura 5.18c</b> – Violação de topologia corrigida, menor tensão aplicada .....	125
<b>Figura 5.18d</b> – Aplicação de uma tensão inferior, resolução da violação de topologia .....	125
<b>Figura 5.19</b> – Contextualização da generalização das curvas de nível com as linhas de água .....	126
<b>Figura 5.20</b> – Resultados da metodologia aplicada a curvas de nível e linhas de água .....	128



## Lista de acrónimos

25k	- Escala 1/25.000
50k	- Escala 1/50.000
250k	- Escala 1/250.000
500k	- Escala 1/500.000
AGENT	- Automatic GEneralization New Technology
ATKIS	- Official Authoritative Topographical Cartographic Information System
BDG	- Base de Dados Geográficos
CAD	- Computer Aided Design
CART	- Classification And Regression Trees
C&R	- Árvore de Classificação e Regressão
C5	- Árvore de Decisão recorrendo ao algoritmo C5
CHANGE	- Institute of Cartography - University of HANover - GEneralization Software
DCM	- Digital Cartographic Model
DEM	- Digital Elevation Model
DLM	- Digital Landscape Model
DMA	- Defense Map Agency
DTM	- Digital Terrain Model
EMD	- Decomposição de modo empírico
ESRI	- Environmental Systems Research Institute
EUA	- Estados Unidos da América
GAEL	- Generalisation based on Agents and Elasticity
GDB	- Geographic Data Base
GRID	- Grelha Regular de pontos
IA	- Inteligência Artificial
ICA	- International Cartographic Association
IGeoE	- Instituto Geográfico do Exército
IGN	- Institut Geographique National
KBS	- Knowledge Based System
KRA	- Knowledge Representation and Abstraction Model
LOD	- Level Of Detail
LTU	- Logic Threshold Unit
M888	- Série Cartográfica do IGeoE, escala 1/25.000
M782	- Série Cartográfica do IGeoE, escala 1/50.000
MDT	- Modelo Digital do Terreno
NHD	- National Hydrography Dataset
NMA	- National Mapping Agency
NN	- Neuronal Network
NOAA	- National Oceanic and Atmospheric Administration
OCL	- Object Constraint Language
Pc	- Percentagem de concordância
PlaGe	- Platform for research in Generalization

RBC	- <i>Raciocínio Baseado em Casos</i>
SCE	- <i>Serviço Cartográfico do Exército</i>
SERVIR	- <i>Sistema de Estações de Referência GPS VIRTuais</i>
SIG	- <i>Sistemas de Informação Geográfica</i>
SGBD	- <i>Sistema de Gestão da Base de Dados</i>
SOM	- <i>Self Organizing Map</i>
TIN	- <i>Triangular Irregular Network</i>
UML	- <i>Unified Modeling Language</i>
USGS	- <i>United States Geological Survey</i>
UTM	- <i>Universal Transverse Mercator</i>
VB	- <i>Visual Basic</i>
VG	- <i>Vértice Geodésico</i>
WFS	- <i>Web Feature Services</i>
WMS	- <i>Web Map Services</i>

# Capítulo 1

## Introdução

---

### 1.1 Introdução

A generalização cartográfica consiste num conjunto de operações realizadas sobre a representação da informação geográfica com o objetivo de melhorar a legibilidade e aumentar a facilidade de compreensão dos dados por parte do utilizador.

A generalização cartográfica desenvolveu-se nos anos 60 do século passado nas Agencias Cartográficas Nacionais (NMA, *National Mapping Agency*) como tarefa integrante de um fluxo de trabalho de produção cartográfica (Ruas, 2002). Nestas cadeias de produção, os dados de trabalho originais são as séries cartográficas de escala superior, com base nas quais se produzem mapas de escala inferior. Este processo decorre da necessidade de produção multi-escala a que as NMAs estão obrigadas exigindo a manutenção de enormes volumes de dados de diferentes resoluções espaciais e temporais de difícil gestão e manutenção. A alternativa a este ciclo de crescente aumento de volume de dados, reside na implementação de sistemas de generalização automática que permitam reduzir a produção e manutenção de uma única série cartográfica de maior exatidão geométrica e semântica. Algumas NMAs, como é o caso do IGeoE (*Instituto Geográfico do Exército*), mantêm duas BDG (*Base de Dados Geográficos*), uma de média escala (25k) e outra de pequena escala (250k), a partir destas produzem as suas séries cartográficas, escalas 25k a 500k, respetivamente. A generalização cartográfica praticada nas NMAs é muito morosa prevalecendo os processos semiautomáticos de edição cartográfica realizados por operadores cartográficos. Ou seja, o computador efetua

tarefas de baixo nível, sequenciamento lógico de operações, ficando a parte de alto nível, cognitiva e decisória do lado do cartógrafo. O cartógrafo recorre desta forma a algoritmos ou sequências de algoritmos, que traduzem os operadores de generalização. Contudo, o resultado final da generalização de uma carta não corresponde a uma simples soma de processos de generalização individual dos elementos que compõem a carta.

Na atualidade estão disponíveis inúmeros algoritmos para generalização, seja em sistemas comerciais como o Dynagen da Intergraph, e Lamps2 v5-2 ou Clarirty 2.0 da Laser-Scan (Mackaness et al. 2007), seja no meio académico e.g. CHANGE do Institute of Cartography – University of Hanover, (Mackaness et al. 2007) seja nas NMAs (Müller et al. 1995b). Em todos é mantida uma abordagem de sequenciamento e parametrização de algoritmos que funcionam por níveis e dados, requerendo a interação do operador e afinação de parâmetros.

Para a automatização do processo de generalização de mapas, é necessário integrar a experiência dos cartógrafos com as operações de generalização nos Sistemas de Informação Geográfica. A aplicação de métodos de inteligência artificial, parece ser uma metodologia promissora para integrar a experiência dos cartógrafos nestes sistemas de generalização. Os métodos de inteligência artificial acomodam uma grande quantidade de interpretação, julgamento e de conhecimento heurístico, que especifica um conjunto de ações a serem executadas para uma dada situação (Müller et al. 1995b; Ladeira, 1997). Isto é feito com a simulação do conhecimento de um especialista humano e respectivo raciocínio, de modo que possa ser formulado nas partes do conhecimento, tipificado por um conjunto de factos e de regras heurísticas. Ou seja, os métodos de inteligência artificial são dispositivos de comunicação entre o conhecimento de um utilizador experiente e um programa de computador a fim de resolver problemas complexos. Estes métodos tentam reduzir o custo e o tempo, simultaneamente aumentando a exatidão, a estabilidade e a consistência.

Pretende-se com a presente investigação dar um contributo efetivo na automatização da generalização cartográfica. O universo da pesquisa é a cartografia de base produzida pelo IGeoE de qualidade cartográfica correspondente à escala 1:25.000, a partir da qual se pretende obter por generalização um produto cartográfico de qualidade 1:50.000. O estudo centra-se na generalização do tema da altimetria, nomeadamente as curvas de nível e entidades cartográficas que com estas se relacionam. Em particular será estudado o problema da contextualização da generalização das curvas de nível com as linhas de água, pontos de cota e

vértices geodésicos. A metodologia resultante deste estudo poderá por analogia ser utilizada noutras escalas, outros temas e englobando diferentes entidades cartográficas.

## **1.2 Objetivos**

Este trabalho tem como objetivo o desenvolvimento e implementação de uma infraestrutura conceptual de generalização cartográfica automática baseada numa estratégia de constrangimento e refinamento iterativo com implementação de soluções multicritério. Pretende-se conceber uma infraestrutura conceptual que automatize o processo de generalização cartográfica com suporte interativo. O desafio é combinar, homogeneizar e desenvolver, as diversas abordagens para generalização de cartografia de base topográfica. Optou-se por centrar o estudo na generalização das curvas de nível e na contextualização com pontos cotados e linhas de água.

Desta forma os objetivos propostos incluem:

- Estudar e propor um algoritmo para generalização de linhas, que possa ser aplicado na generalização cartográfica de curvas de nível.
- Propor um conjunto de atributos classificadores de linhas, para classificação das curvas de nível.
- Estudar uma metodologia baseada em IA (*Inteligência Artificial*) para, recorrendo à classificação das linhas, escolher automaticamente o parâmetro a utilizar na generalização cartográfica das curvas de nível.
- Estudar uma metodologia para detetar e resolver violações de topologia entre curvas de nível, pontos de cota e linhas de água.

## **1.3 Organização da dissertação**

Esta tese está organizada em seis capítulos, sendo este o **capítulo** onde é apresentado o problema a ser estudado, as motivações para o seu estudo e a metodologia proposta.

O segundo capítulo refere-se ao estado da arte em generalização cartográfica, onde se apresentam os principais modelos conceptuais de generalização em SIG (*Sistemas de Informação Geográfica*). Apresentam-se também os operadores e algoritmos que os

traduzem, faz-se ainda uma alusão à classificação, segmentação e caracterização de linhas, como tarefa integrante da estratégia adotada para a automatização da generalização de curvas de nível.

No terceiro capítulo aborda-se a temática do conhecimento em generalização e utilização da IA. Desde a definição de conhecimento, sua formalização e exploração, os mecanismos de abstração, aquisição de conhecimento e respetiva representação. É ainda neste capítulo que são apresentados os métodos de IA utilizados nesta dissertação. No quarto capítulo caracterizam-se as séries cartográficas envolvidas, aborda-se o tema da generalização modelo, classificação e seleção das linhas de água. Apresentam-se também neste capítulo algoritmos de generalização de linhas, incluindo o algoritmo proposto nesta dissertação. No quinto capítulo apresenta-se a metodologia e sua aplicação às séries cartográficas do IGeoE. No sexto capítulo discutem-se os resultados, apresentam-se as conclusões e perspectivas futuras.

# Capítulo 2

## Generalização cartográfica

---

### 2.1 Generalização cartográfica

A investigação na área da generalização cartográfica evoluiu de uma fase inicial centrada no desenvolvimento de algoritmos, para uma fase recente, mais focada na formalização e representação do conhecimento cartográfico através de modelos conceptuais abrangentes e na utilização de sistemas periciais. Este segundo capítulo pretende dar a conhecer o trabalho efetuado em generalização, a sua evolução e principais desenvolvimentos relevantes para esta tese. Começa-se com uma perspetiva evolutiva da generalização cartográfica manual até à generalização digital, pela utilidade e/ou necessidade da generalização, modelos propostos para generalização, generalização de linhas, sua segmentação e classificação, operadores de generalização e algoritmos.

A finalidade da generalização cartográfica é produzir um mapa, balanceando requisitos de exatidão, conteúdo da informação e legibilidade. Na construção de um mapa os objetos da superfície terrestre a cartografar, são representados de forma reduzida, para que o observador tenha uma boa perceção da realidade que o mapa pretende transmitir. Para que isto seja possível, é necessário selecionar os objetos a representar, simplificar formas e estruturas e respeitar critérios de importância relativa (Robinson, 1960; Taillandier et al. 2009).

A generalização cartográfica é um processo multidisciplinar e complexo, para o qual têm sido propostas soluções pontuais para problemas específicos, assentando essencialmente no desenvolvimento de algoritmos para a implementação de operadores de generalização. A

dificuldade na automatização da generalização cartográfica decorre da aquisição e utilização do conhecimento mais profundo, detido por cartógrafos, que recorrem a um raciocínio mais complexo e habilidade ou capacidade para tirar conclusões baseadas no contexto geográfico, prioridades, padrão ou finalidade do mapa. É então necessário ter uma visão holística da área do mapa e ter em consideração as prioridades e relacionamentos entre os objetos representados, relativos à sua finalidade.

## 2.2 Da generalização cartográfica manual à digital

O período da generalização digital, teve o seu início durante os anos 60. O termo “*generalização digital*” era usado frequentemente para enfatizar a transição de uma era manual para a era informática digital (Shea e McMaster, 1989; Buttenfield e McMaster, 1991). Um dos resultados mais conhecidos desse período é a lei (empírica), apresentada por Töpfer e Pillewizer (1966), de acordo com a qual se poderia calcular o número de símbolos a representar num mapa em diferentes escalas. Nos seus estudos sobre generalização cartográfica, Töpfer (1966) apresentou o seu “princípio da seleção”, que estabelecia que a quantidade de informação que pode ser representada por unidade de área, decresce de acordo com uma progressão geométrica. Na sua forma mais simples, o princípio de Töpfer indicava a quantidade de objetos que deveriam ser representados num mapa. Töpfer expressou uma relação entre a escala do mapa e o número de itens a representar pela seguinte equação:

$$n_c = n_s \sqrt{\frac{S_c}{S_s}} \quad (2.1)$$

Onde:

$n_c$  = número de itens num mapa compilado na escala  $S_c$ ;

$n_s$  = número de itens num mapa original na escala  $S_s$ ;

Como resultado da aplicação desta fórmula obtemos o número de símbolos a serem representados. No entanto, não revela quais destes símbolos devem ser selecionados e não tem em consideração a variação local da densidade dos fenómenos representados. Esta lei da seleção não foi especificamente apresentada para a generalização automática, contudo, foram efetuados esforços para a formulação matemática de problemas de generalização. De qualquer modo, o princípio facilita o trabalho de compilação, pois fornece o número aproximado de



objetos que podem ser representados no mapa generalizado. Além disso, permite uniformizar a generalização efetuada por diferentes cartógrafos que trabalham em folhas adjacentes da mesma série cartográfica (Töpfer e Pillewizer, 1966). Foi neste primeiro período da generalização digital, que Douglas e Peucker (1973) desenvolveram um dos algoritmos mais usados e relevantes de simplificação de linhas.

Durante o segundo período da generalização digital, no final dos anos 70 e princípio dos anos 80, a eficiência do algoritmo de Douglas e Peucker foi enfatizada (Buttenfield e McMaster, 1991; Li, 2007) tendo sido feitas propostas de alterações ao algoritmo. Neste período o maior esforço de investigação sobre generalização centrou-se em aspetos muito específicos, não olhando para o problema de uma forma abrangente (Weibel, 1991), tal como a simplificação de linhas e seleção de objetos de acordo com um determinado valor. Abordagens mais complexas incluíram os primeiros estudos em deslocamento automático. Christ (1978) estudou o deslocamento automático de símbolos pontuais e lineares, Lichtner (1979) desenvolveu algoritmos para o deslocamento automático considerando as posições relativas dos objetos numa determinada escala.

Nos anos 1980 a investigação e desenvolvimento foi direcionada para os aspetos da modelação conceptual da generalização. Foram propostos dois métodos de generalização, conceptual e estrutural (Bertin, 1983). A generalização conceptual é uma modalidade de generalização cartográfica em que a forma de representar os fenómenos é alterada, de modo a adequá-la a um novo conceito, ou a uma alteração do nível de abstração, e.g. o colapso é uma forma de generalização conceptual. A generalização estrutural é uma modalidade de generalização cartográfica em que a forma de representação e implantação dos fenómenos é conservada, mas a sua estrutura original é simplificada, e.g. a agregação é uma forma de generalização estrutural (Gaspar, 2008).

Nos anos 90 os investigadores continuaram a propor algoritmos de generalização, traduzindo os respetivos operadores. Visvalingham e Whyatt (1993) desenvolveram um algoritmo para melhorar a distorção da forma. Outras abordagens englobaram o desenvolvimento de algoritmos de redução de pontos auto-adaptáveis. Neste caso há um pré-processamento e, a linha é inicialmente segmentada, sendo posteriormente aplicado o algoritmo aos respetivos segmentos, com diferentes critérios e parâmetros e.g. (Aoyama e Kawagoe, 1991; Sato, 1992).

Wang e Muller (1998) apresentaram um algoritmo para segmentação de estradas. Muitos autores utilizaram Inteligência Artificial para classificar estes segmentos, Balboa et al. (2008) utilizaram uma rede neuronal para classificar estradas, Plazanet et al. (1998) utilizaram uma árvore de decisão, recorrendo ao algoritmo ID3, para classificação de estradas utilizando como input a sinuosidade e o grau de conflito devido à simbolização. Um outro grupo de algoritmos inclui a suavização de linhas. Foram utilizadas várias técnicas como a convolução Gaussiana, transformação de Fourier (Boutoura, 1989; Plazanet et al. 1996), transformação de wavelet e.g. (Balboa e López, 2000), snakes (Burghardt e Meier, 1997; Steiniger e Meier, 2004; Burghardt, 2005; Bader e Barrault, 2000), decomposição de modo empírico (EMD) (Li et al. 2004a) e fractais (Müller, 1986; Buttenfield, 1989). Contudo, os parâmetros usados nos algoritmos não estão relacionados diretamente com os níveis de detalhe dos mapas e respetivas escalas. Ou seja, a parametrização dos algoritmos ou operadores é efetuada a priori independentemente da razão entre as escalas dos mapas.

Surgiram também neste período tentativas de resolver o problema da generalização como soma de vários pequenos problemas. Neste caso estão os trabalhos desenvolvidos para deslocar edifícios, como entidades areais (e.g. Ruas, 1998; Taillandier e Taillandier, 2012), tendo sido desenvolvidos métodos como o campo do deslocamento (Ai, 2004), o método dos elementos finitos (Højholt, 2000), o método de elementos finitos com suporte dúctil (Bader et al. 2005) e ajustamento por mínimos quadrados (Harrie, 1999; Harrie e Sarjakoski, 2000; Sarjakoski e Kilpeläinen, 1999; Sester, 2000). Outros algoritmos foram desenvolvidos para a tipificação de edifícios (Regnault, 1998, 2001a, 2001b; Basaraner e Selcuk, 2008; Li et al. 2004b). Foi também neste período que se iniciou a discussão em torno do conhecimento envolvido na generalização digital de mapas. Este assunto foi discutido por muitos investigadores tendo sido identificados diversos tipos de conhecimento em generalização. Este assunto merecerá um desenvolvimento aprofundado no próximo capítulo.

No final dos anos 1990 surgiram as primeiras propostas de introdução de confinamentos no processo de generalização cartográfica, são exemplos: (Beard, 1991; Ruas e Lagrange, 1995; Ruas, 1998; Weibel e Dutton, 1999; Peter e Weibel, 1999; Galanda, 2003; Steiniger e Weibel, 2005). Peter e Weibel (1999) classificam os confinamentos de generalização em quatro tipos, confinamentos gráficos, topológicos, estruturais e de forma, de acordo com as suas funções para a generalização categórica do mapa. Vários tipos de confinamentos foram usados por exemplo, para o deslocamento de edifícios (Harrie, 1999; Sarjakoski e Kilpeläinen, 1999;

Sester, 2000). De facto, alguns confinamentos podem ser expressos em termos de relações espaciais, isto é relações topológicas, métricas ou de ordem.

Outra linha de investigação incluiu o enriquecimento de bases de dados (Plazanet et al. 1998; Plazanet, 1996; Neun et al, 2004), para a aplicação de agentes em generalização (Lamy et al, 1999; Galanda e Weibel, 2002; Gaffuri, 2007; 2008) e, na sua generalização para transmissão de dados na internet (Foerster, 2010; Neun e Burghardt, 2005).

Para Li (2007) no próximo período de desenvolvimento em generalização, serão usados os algoritmos para a generalização de uma classe de entidades, isto é, relevo, entidades hidrográficas, edifícios, rede de transporte, etc. Segundo este autor, os sistemas semiautomáticos ainda estarão disponíveis por mais alguns anos. Sistemas mais inteligentes só serão desenvolvidos depois da generalização ao nível da classe estar completamente estudada. A generalização cartográfica digital tornar-se-á cada vez mais importante em multi-escala, modelação e representação, com a utilização generalizada dos SIG que integram dados multi-escala e de várias origens, como BDG distribuídas.

## **2.3 Modelos conceptuais de generalização**

No âmbito da generalização cartográfica foram desenvolvidos modelos conceptuais para generalização da informação com o objetivo de sistematizar e modelar o processo de acordo com um conjunto de regras que preservem as características geométricas e topológicas dos objetos.

### **2.3.1 Modelo de Ratajski**

Ratajski (1967) apresentou um dos primeiros modelos formais da generalização e identificou dois tipos fundamentais de processos da generalização: quantitativo e qualitativo. A generalização quantitativa consiste numa redução gradual no conteúdo do mapa dependendo da mudança de escala, enquanto a generalização qualitativa resultaria da transformação de formas elementares da simbolização para formas mais abstratas (McMaster, 1991). Na sua estrutura, Ratajski descreve as primeiras etapas da generalização com a eliminação dos objetos secundários, preservando as características do mapa. Este limite de representação só poderia ser superado se uma determinada soma de elementos pudesse ser substituída por métodos cartográficos novos. Ratajski (1967) definiu estes limites como pontos de

generalização. A mudança de capacidade de um mapa<sup>1</sup> pode ser representada por um triângulo, ver figura 2.1, onde a base ilustra a capacidade máxima e a parte superior ilustra os limites de capacidade do mapa. Quando a capacidade do mapa está perto da parte superior do triângulo, deve ser aplicado um método cartográfico novo, a fim de iniciar um novo ciclo de generalização.

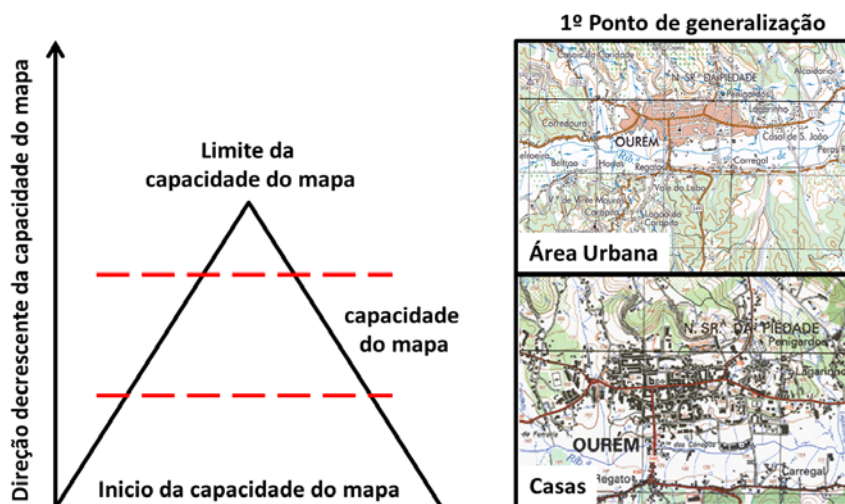


Figura 2.1 – Modelo da generalização proposto por Ratajski (adaptado de Cecconi, 2003)

A figura 2.1 ilustra o modelo da generalização proposto por Ratajski. Do lado esquerdo da imagem temos representado o triângulo de mudança de capacidade do mapa, do lado direito podemos ver um exemplo do 1º ponto de generalização. Perto da capacidade limite, as casas devem ser substituídas por uma área, representando a área urbana.

Durante os finais dos anos 80 e início dos 90, houve uma mudança de direção nos desenvolvimentos, sendo colocada uma maior ênfase na modelação de dados, envolvida nas tecnologias de base de dados. Este período conduziu ao duplo conceito da generalização, em generalização cartográfica e generalização do modelo.

### 2.3.2 Modelo de Brassel e Weibel

Brassel e Weibel (1988) propuseram uma estrutura conceptual para a generalização automática, diferenciando generalização estatística e cartográfica. A generalização estatística

<sup>1</sup> A **capacidade do mapa** - depende da escala de representação, da densidade dos objetos representados, da simbologia utilizada, entre outros fatores. Se a escala diminui ou se a densidade de objetos aumenta, então a capacidade de representação dos objetos de uma forma legível diminui.

foi descrita como um processo analítico que trata da redução do conteúdo da informação da base de dados, sob um controlo estatístico. A generalização cartográfica foi considerada como a modelação espacial para uma comunicação visual, visando a modificação da estrutura local.

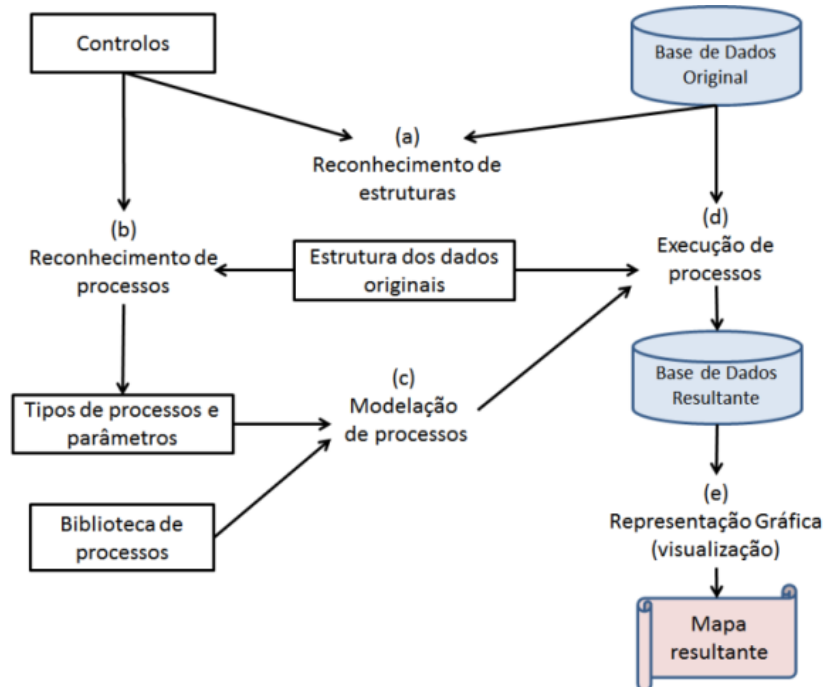


Figura 2.2 - Modelo conceitual de Brassel e Weibel, 1988, para a generalização, adaptada de (Mackaness et al. 2007).

Este modelo apresenta uma sequência de etapas que permitem escolher os procedimentos de generalização de maneira adaptativa. Há uma ênfase na generalização cartográfica como parte da modelação espacial e em modelos de dados alternativos, como condição para uma automatização satisfatória do processo de generalização. Foram identificados cinco processos distintos de generalização no domínio digital: reconhecimento de estruturas, reconhecimento de processos, modelação de processos, execução de processos e visualização de dados, ver figura 2.2.

A fase de reconhecimento de estruturas identifica os objetos cartográficos específicos, ou agregação de objetos, bem como as relações espaciais e estabelece medidas de importância relativa entre os objetos. Esta medida é controlada pela qualidade da base de dados original e pela escala do mapa a ser produzido.

A fase de reconhecimento de processos tem o objetivo de definir o processo de generalização, identifica os operadores de generalização necessários, bem como o tipo de modificação dos dados e a seleção de parâmetros. Esta atividade determina o que será feito com os dados

originais, quais os tipos de conflito identificados e resolvidos e, quais os tipos de objetos e estruturas que serão considerados no mapa a ser produzido.

A fase da modelação de processos compila regras e procedimentos, que serão usados na etapa seguinte de execução de processos, para generalizar a base de dados original. É a etapa onde os algoritmos, especificados para as operações identificadas, serão aplicados utilizando regras e parâmetros extraídos da biblioteca de processos, onde estão armazenadas as regras e procedimentos para a generalização. A fase da execução de processos é a etapa onde o processo de generalização é realizado. Por último a visualização dos dados é a etapa onde a base dos dados generalizada é transformada no mapa desejado.

Os elementos fundamentais deste modelo são as estruturas da informação espacial e os processos que podem interagir com tais estruturas. Há uma ênfase em generalização cartográfica como parte da modelação espacial e em modelos de dados alternativos como condição para uma automatização satisfatória do processo de generalização. Vianna (1997), acrescenta que o desenvolvimento da componente biblioteca de processos é uma etapa importante na construção de sistemas periciais para generalização cartográfica em ambiente digital, uma vez que inclui as operações, o conhecimento ou regras para generalização e os valores de tolerância.

### **2.3.3 Modelo de McMaster e Shea**

McMaster e Shea (1992) descreveram um modelo conceptual abrangente com o propósito de discutir a filosofia da generalização no domínio digital. O processo da generalização em ambiente digital, foi dividido em três componentes. O modelo é baseado em três perguntas que pretendem responder o “porquê” generalizar, sendo este o primeiro componente deste modelo, de condições que indicam o “quando” generalizar (segundo componente) e de transformações espaciais e de atributos que definem o “como” generalizar, ver figura 2.3.

Este modelo impõe uma abordagem tipicamente algorítmica, pois baseia-se no uso de transformações espaciais e de atributos para materializar as ações de generalização. O problema é que a aplicação de uma sequência de transformações envolve parâmetros que se relacionam mutuamente e não são completamente independentes. Na verdade, é muito difícil controlar uma transformação sem interferir nas outras.

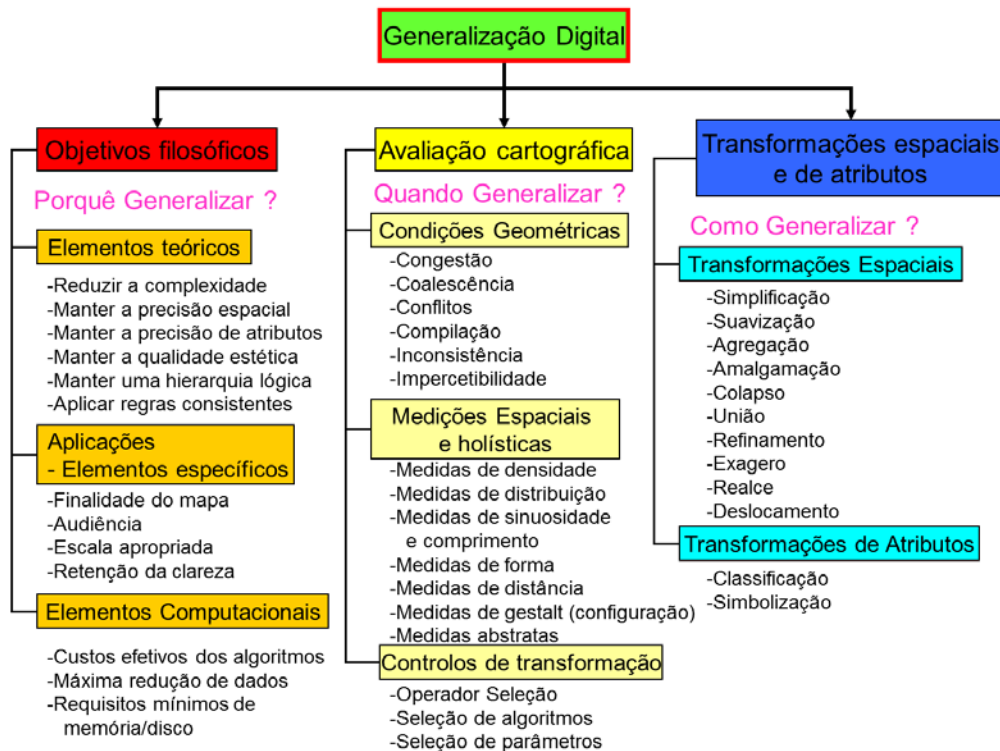


Figura 2.3 - Modelo conceitual de McMaster e Shea (1992) para a generalização em ambiente digital, adaptada de (Mackaness et al. 2007).

O segundo componente é a avaliação cartométrica, onde é necessário definir “quando é necessária a generalização”, a fim de identificar as condições específicas onde a generalização é exigida. Embora muitas condições importantes possam ser identificadas, as condições fundamentais apresentadas por (McMaster e Shea, 1992) incluem: Congestionamento, coalescência, imperceptibilidade, complexidade, inconsistência e conflito. Apesar de que muitos problemas em generalização requererem o desenvolvimento e implementação de medidas matemáticas, estatísticas ou geométricas, McMaster e Shea (1992) apresentam uma classificação geral de medidas incluindo as seguintes classes: densidade, distribuição, comprimento e sinuosidade, forma, distância e gestalt<sup>2</sup>.

Neste modelo, o terceiro e último componente são as operações fundamentais ou “como generalizar”. Este componente é o responsável pela execução do processo de generalização cartográfica, através de transformações espaciais e de atributos, submetidas aos dados digitais. Em muitos casos estas transformações não são independentes, existindo um relacionamento

<sup>2</sup> **gestalt** - Palavra alemã que se pode traduzir por “conformação”, “estrutura”, “organização”...realidade psicológica onde existe alguma coisa mais do que a simples soma dos elementos que a constituem. (Grande Enciclopédia Portuguesa Brasileira)

entre si. As transformações espaciais são realizadas pelos operadores que alteram a representação dos dados digitais do ponto de vista geográfico ou topológico. Foram identificados dez operadores: simplificação, suavização, agregação, amalgamação, união, colapso, refinamento, exagero, realce e deslocamento. As transformações de atributos manipulam as características secundárias dos elementos. Foram identificadas duas transformações de atributos: classificação e simbolização. Embora este modelo abrangente, tenha influenciado indubitavelmente as pesquisas mais recentes em generalização, apresenta a limitação de não considerar correlações existentes entre os parâmetros envolvidos nas várias transformações. Desde os anos 90 que estas considerações têm sido largamente discutidas, principalmente os aspetos relativos ao porquê-quando-como.

### **2.3.4 Modelo de Grünreich**

O que anteriormente foi chamado generalização estatística, na estrutura de Brassel e Weibel (1988), esteve na origem do que mais tarde foi apelidado como generalização conceptual ou do modelo. Foi necessário efetuar a distinção entre, por um lado, os processos da generalização associados às BDG e por outro, os processos necessários para a representação cartográfica final dos mapas de saída.

Como indicado por Kilpeläinen (1997) e Weibel e Dutton (1999), a distinção entre a generalização do modelo e a generalização cartográfica eram já evidentes nas primeiras obras de Grünreich (1985) que propôs a criação de uma representação primária do mundo real, denominada DLM (*Digital Landscape Model*). A partir desta representação, bastante detalhada, seria possível derivar representações secundárias, específicas para uma aplicação ou grupo de aplicações com necessidades semelhantes em termos de detalhe. Estas representações secundárias foram denominadas DCM (*Digital Cartographic Models*). O objetivo é que o DLM seja uma Base de Dados com bastante detalhe, criado sem a preocupação de refletir as necessidades cartográficas, mas sim associado a um parâmetro básico de precisão. Este parâmetro está relacionado obviamente à fonte de dados e ao processo de aquisição e reflete o nível de detalhe do DLM. Um DLM pode ser trabalhado, usando algoritmos específicos, para produzir outro DLM, menos detalhado, ou para produzir um DCM adequado para uma saída e utilização numa determinada escala. O processo de transformação DLM-DLM foi denominado generalização conceptual, ou do modelo,



enquanto a transformação DLM-DCM foi denominada generalização cartográfica, ver figura 2.4.

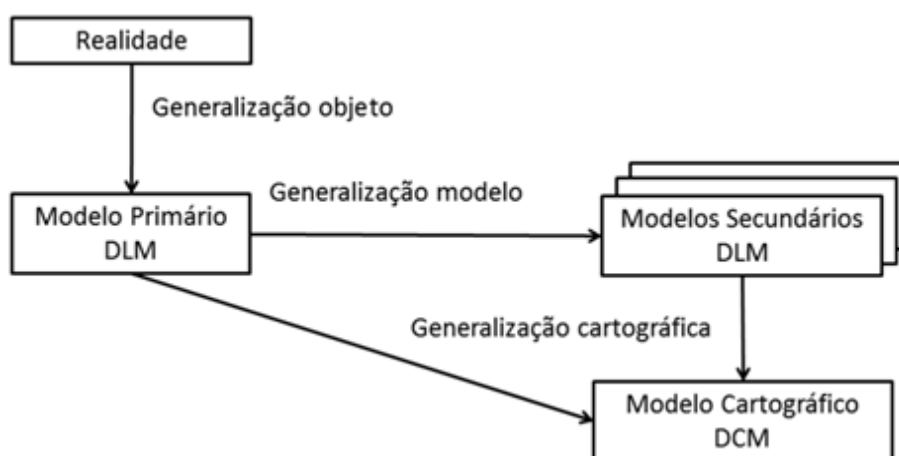


Figura 2.4 – Modelo de generalização de Grünreich (1992), adaptado de Foerster et al. (2010)

Este modelo foi usado no desenvolvimento do ATKIS (*Official Authoritative Topographical Cartographic Information System*) na Alemanha, durante os finais dos anos 80 (Grünreich 1992, 1995a, 1995b). Os DLMs foram adquiridos pela generalização objeto a partir de um estado do mundo real e resultando os DCMs da generalização cartográfica. A finalidade da generalização modelo era criar uma base de dados modificada, usando a base de dados geográfica primária como fonte de entrada.

Durante a fase da generalização conceptual, os aspetos relativos à exposição gráfica não foram considerados. A generalização modelo foi considerada como uma etapa de pré-processamento antes da visualização através da generalização cartográfica e assim a generalização modelo (conceptual) não envolveu nenhum componente artístico ou intuitivo. A generalização modelo fornece os dados geográficos para funções de análise, visto que a generalização cartográfica é executada para a visualização de produtos cartográficos. Neste contexto, a generalização cartográfica envolve a redução, ampliação e modificação do simbolismo gráfico num mapa a ser executado, de modo a aumentar a eficácia da comunicação cartográfica. Kilpeläinen (1992a, 1992b, 2001) enfatizou que a semântica dos dados espaciais pode mudar na generalização do modelo, e.g. a criação de áreas de blocos de edifícios em cidade pela agregação dos respetivos edifícios.

Grünreich (1995b) sugeriu que generalização no domínio digital deveria basear-se num modelo cartográfico que descrevesse completamente o processo de comunicação da

geoinformação através de modelos específicos das disciplinas envolvidas, de modelos projetados para comunicação visual e de modelos resultantes de estudos cognitivos. Postulou ainda, que o conhecimento sobre generalização precisa ser adquirido e formalizado para que seja possível desenvolver soluções computacionais eficientes.

Este modelo envolve os seguintes conceitos de generalização, conforme figura 2.4:

**Generalização objeto** - com o levantamento de dados, os objetos são abstraídos do mundo real, mas com uma quantidade finita de observações e medidas. Os objetos são selecionados e simplificados, isto é generalizados.

**A generalização modelo** - executa uma redução controlada ‘dos dados’ (na base de dados), para gerar uma quantidade menor de objetos para o nível de detalhe desejado.

**A generalização cartográfica** - a aparência visual dos objetos é mudada, e.g. simbolização de uma igreja, ou deslocamento de objetos, de modo que estes não se sobreponham numa escala menor, melhorando dessa forma a legibilidade do mapa.

### 2.3.5 Modelo AGENT

O projeto AGENT (*Automatic GEneralization New Technology*) financiado pela UE, iniciou-se em 1997 e englobava as seguintes entidades, Institut Géographique National (França), Laser-Scan Ltd (Reino Unido), Institut National Polytechnique de Grenoble (França), University of Edinburgh (Reino Unido) e University of Zurich (Suíça). Neste projeto a generalização cartográfica foi considerada como um problema de optimização. Foram desenvolvidas algumas abordagens de optimização para a generalização cartográfica, como o método das *snakes* (Bader e Barrault, 2000), *elastic beams* (Bader e Barrault, 2001) e ajustamento por mínimos quadrados (Sarjakoski e Kilpeläinen, 1999). Neste grupo podemos incluir o projeto AGENT, proposto por Ruas e Plazanet (1996). A proposta assentava num modelo controlado por um conjunto de constrangimentos. O objetivo do projeto AGENT assenta na modelação da natureza holística da generalização, recorrendo a tecnologia multiagente (Maudet et al. 2014). Em vez de usar um plano centralizado no processo da generalização, usa constrangimentos locais, regionais e globais, que são armazenados nos agentes.

A abordagem baseada em agentes aplica o paradigma orientado a objetos. Um agente é qualquer coisa que pode ser visto como algo que percebe o seu ambiente através de sensores e age nesse ambiente através de mecanismos (Russel e Norvig, 2003). Um agente pode atuar como um objeto que tem um objetivo e atua autonomamente de modo a atingir o seu objetivo graças às capacidades de percepção, deliberação, ação e possibilidade de comunicação com outros agentes. Estes agentes são os objetos cartográficos (micro agentes), grupo de objetos (meso agentes) ou as classes de objetos (macro agentes) que podem comunicar e afetar outros agentes, ver figura 2.5.

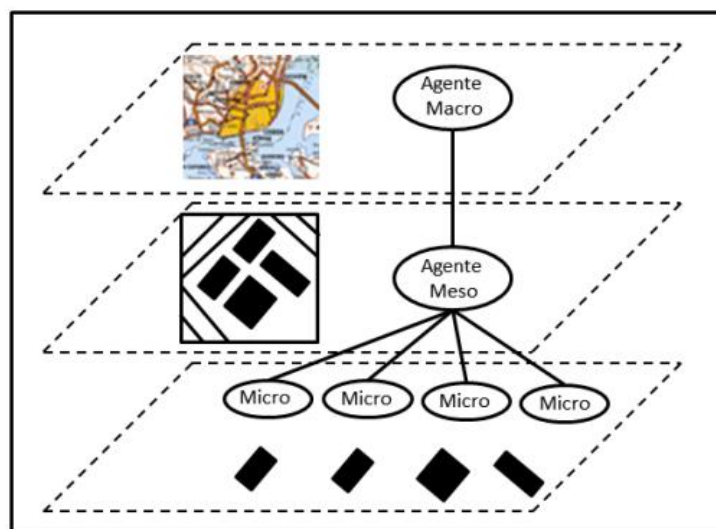


Figura 2.5 – Modelo de generalização AGENT, relação entre agentes micro, meso e macro

Um único objeto cartográfico é controlado por um micro agente e um grupo de objetos cartográficos, i.e. um grupo de micro agentes, é controlado por um meso-agente. O macro agente controla todos os objetos cartográficos no display do mapa, para maximizar objetivos de otimização globais. Um agente macro pode ser considerado de um tipo especial de meso-agente, pois também controla um conjunto de objetos.

O objetivo de cada um destes agentes é melhorar a situação no que diz respeito aos conjuntos de confinamentos, ou seja, tenta maximizar o propósito de objetivos de otimização. Cada confinamento fornece 'um valor do objetivo' que cada agente tenta alcançar generalizando-se a si próprio e em comunicação e interação com outros agentes. O estado de um agente é caracterizado por todos os seus confinamentos que integram os seus valores e importância para estabelecer o seu próprio estado. Quando o seu estado não estiver conforme, o agente

propõe uma ação de generalização a fim de remediar uma violação do confinamento. O objetivo é melhorar a busca para a melhor sequência, os agentes classificam as ações disponíveis no que diz respeito à prioridade, severidade e classificação dos confinamentos. O agente então provoca a primeira ação e reavalia o seu próprio estado. Dependendo do sucesso do plano, ver figura 2.6, pode retroceder e iniciar o plano seguinte, parando se for alcançado um estado perfeito, ou iniciando um novo ciclo se o estado melhorou, até não haver mais nenhum plano adicional.

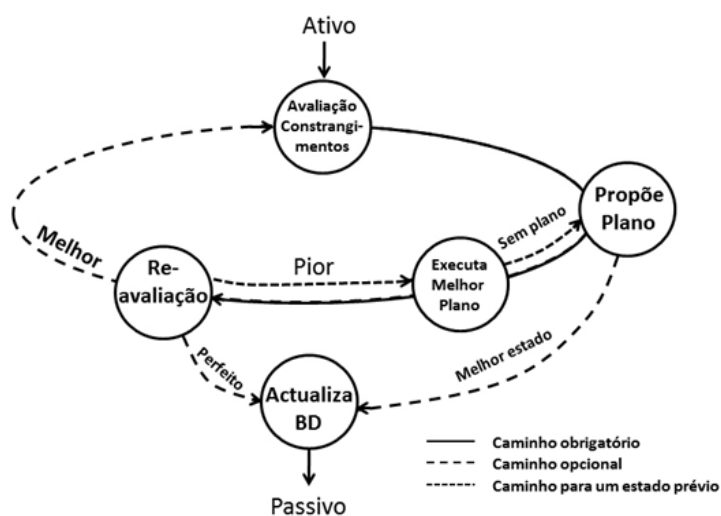


Figura 2.6 – Modelo de generalização AGENT, proposto por (Galanda e Weibel, 2002)

Os resultados do projeto AGENT foram implementados no software Clarity. Uma descrição mais detalhada do projeto AGENT pode ser vista em (Lamy et al, 1999; Ruas, 2000a; 2000b; Barrault et al. 2001; Bader et al. 1999). Apesar deste sistema ser um grande avanço na automatização da generalização, é necessário ter em conta que a interação com os objetos ainda é feita com recurso a algoritmos e estes necessitam de parâmetros para limitar a sua ação. Além disso a parte cognitiva da questão é ainda mais complexa e subjetiva.

## 2.4 Operadores e algoritmos de generalização

A base para qualquer sistema automático de generalização cartográfica assenta em operadores da generalização. Um operador de generalização é, segundo Smith et al. (1999), “*toda a função discreta de generalização que um cartógrafo pode usar para alterar o detalhe ou o conteúdo do todo ou uma parte dos dados do mapa*”. Alguns destes exemplos são a

suavização de uma linha para atenuar o seu detalhe ou a agregação de duas ou mais áreas ou polígonos numa única área.

Numa cadeia de produção cartográfica e para cada tipo de mapa produzido, há conjuntos típicos de operadores usados. Um sistema para executar a generalização cartográfica deve fornecer este conjunto típico de operadores, mas deve permitir igualmente a adição de mais operadores à medida que o cartógrafo aprende sobre o processo da generalização no mundo digital. Há diversos operadores de generalização bem documentados podendo fornecer uma boa base para um sistema de generalização.

Rieger e Coulson (1993) realizaram uma investigação entre um grupo de cartógrafos que executam a generalização interativa e chegaram à conclusão que a classificação de operadores da generalização difere, dependendo de cada cartógrafo. Na bibliografia há diversas classificações e diferentes operadores, pelo que Foerster et al. (2007) propõem uma formalização dos operadores, dividindo em operadores para a generalização modelo e operadores para a generalização cartográfica. O modelo AGENT por exemplo, agrupa os operadores em: transformação de atributos, considerando a classificação dividida em seleção e agregações temáticas e transformações espaciais, agrupando os operadores que atuam em objetos individuais, individuais ou conjunto de objetos e conjuntos de objetos, aqui encontram-se a simplificação, o colapso, o exagero, a amalgamação entre outros, ver (Bader et al. 1999).

De seguida apresenta-se um conjunto de operadores, resultando das diferentes descrições disponíveis, assim como os respetivos nomes usados na literatura. Estes operadores são apresentados sem qualquer relação com algum sistema conceptual de generalização.

### **2.4.1 Classificação**

Um dos aspetos mais importantes do processo de generalização é chamado classificação. Esta transformação está relacionada com o agrupar de objetos, que partilham características idênticas ou semelhantes, em categorias de entidades, ver fig 2.7. Este processo pode ajudar a reduzir drasticamente o número de entidades individuais a representar e, assim, reduzir a complexidade de um mapa. A classificação significa que muitos objetos individuais são agrupados numa classe que representa os seus atributos comuns ou cobertura dominante. No caso da cobertura dominante, a natureza original dos pequenos objetos será mudada. Esta

operação de mudança da natureza é definida também como um tipo de amalgamação. A classificação ajuda a organizar os objetos ou fenómenos nos grupos que são representados pela mesma simbologia. Os objetos ou os fenómenos individuais que são agrupados de tal forma estão, por outras palavras, a ser generalizados, e.g. simbolizar uma área como zona agrícola.

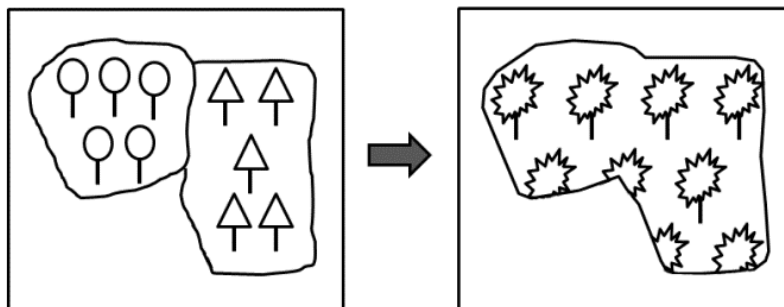


Figura 2.7 – O operador classificação

### 2.4.2 Simbolização

O processo de generalização conhecido como simbolização, figura 2.8, atribui vários tipos de símbolos para refletir a sumarização resultante da classificação em entidades significativas. Estes símbolos devem ser escolhidos com cuidado, de forma a que contribuam para mapas mais legíveis. Os cartógrafos podem usar estes para simbolizar um conceito, uma série de factos, ou o carácter de uma distribuição geográfica.

O processo de simbolização pretende achar boas representações para fenómenos do mundo real ajustando sistematicamente as variáveis visuais. A simbolização pode também requerer a generalização cognitiva, isto é, os cartógrafos podem mudar a dimensão de uma entidade, ou podem mudar a escala da medida de valores do atributo de uma entidade. Esta tarefa implica frequentemente uma mudança da dimensão geométrica, isto é colapso da área para linha, da área para ponto, etc.

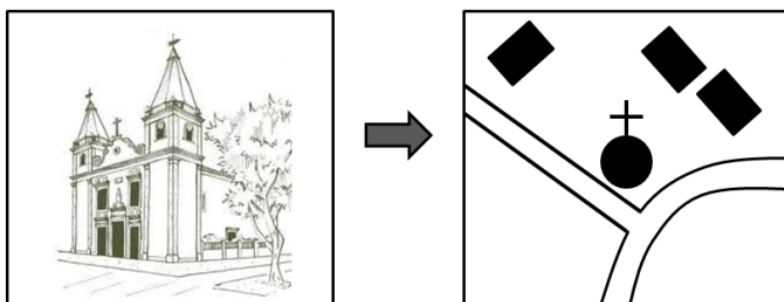


Figura 2.8 – O operador simbolização

### 2.4.3 Agregação

A agregação é uma transformação na qual se agregam pontos da mesma classe de entidades, em entidades de classes de uma ordem mais elevada e são simbolizados como tal. Após as entidades serem generalizadas obtém-se uma melhor clareza na representação. A agregação permite combinar os objetos através da estrutura hierárquica quando a escala se torna menor, por exemplo: os edifícios de uma cidade podem ser agrupados para dar forma a uma mancha urbana, figura 2.9. A generalização semântica ocorre normalmente, antes da generalização geométrica.

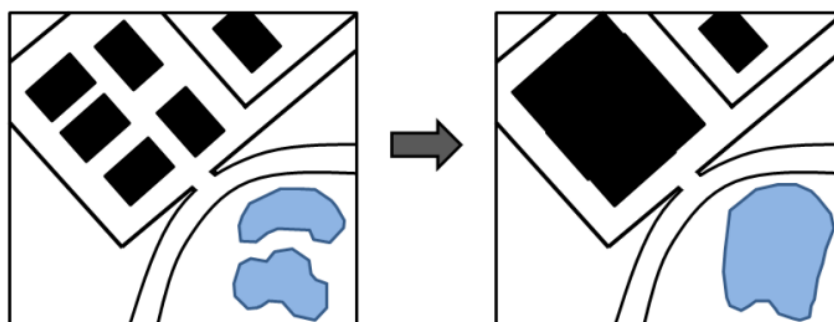


Figura 2.9 – O operador agregação

### 2.4.4 Seleção

A seleção é um processo no qual os objetos que não são pertinentes no mapa generalizado são apagados. Há opiniões diferentes sobre a seleção de objetos e atributos a serem apresentados num mapa. Há autores que consideram que a seleção não é conceptualmente uma parte do processo de generalização, consideram que é um passo de pré-processamento necessário em ambas as transformações, de espaço e de atributos. Outros consideram a seleção como um processo de generalização. Não importa como é categorizado, o processo de seleção tem que ser executado ao criar um mapa e as escolhas que são feitas afetarão o mapa resultante. Na literatura o operador de seleção pode ser também chamado o operador de eliminação. A seleção é aplicada geralmente com o objetivo de escolher as entidades cartográficas, baseadas no seu significado relativo na área abrangida pelo mapa, tal como o significado administrativo, a conveniência de tráfego, frequência da ocorrência, tamanho, etc.

Normalmente os objetos no nível mais baixo da hierarquia são aqueles que são removidos, figura 2.10. No entanto, estes critérios simples do valor relativo podem não ser os mais

corretos se outros critérios, tais como o grau de isolamento forem também avaliados, e.g. um oásis num deserto. Por esta razão o processo de eliminação é a maioria das vezes supervisionado, ou alternativamente feito de forma manual ou semiautomática, mas não inteiramente automática.

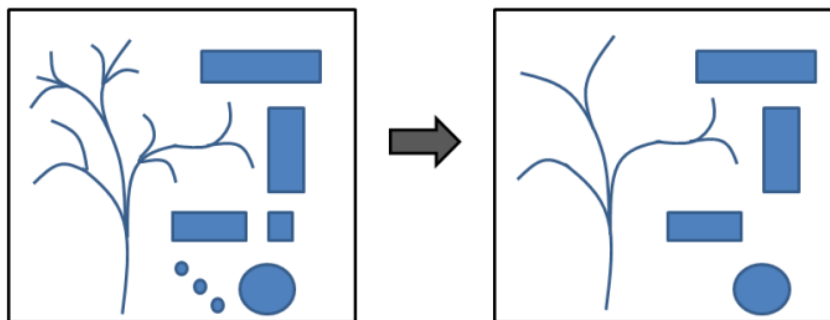


Figura 2.10 – O operador seleção

### **2.4.5 Simplificação**

A simplificação é uma combinação de muitas operações incluindo a eliminação de pequenos objetos do mapa, da redução de pontos numa linha ou superfície, suavização de linhas ou limite de áreas. A simplificação pode ainda incluir ajustamento da posição de cada ponto de uma linha de acordo com a posição dos seus pontos circunvizinhos, a fim de diminuir a angularidade e a preservação de características geométricas, tais como a dimensão fractal de uma linha ou dos cantos retangulares de um edifício.

A representação deve ser tão precisa e detalhada quanto possível. Quando uma entidade é representada num mapa, o objetivo é achar o equilíbrio certo entre minimizar o número de pontos dos dados, usados para exibir uma entidade e representar as características necessárias da entidade. O número excessivo de pontos de dados capturados na fase de digitalização deve ser reduzido selecionando um subconjunto dos pontos originais, retendo os pontos considerados representativos da entidade. Esta transformação espacial é chamada simplificação, ver fig 2.11. O número de pontos representativos da linha são reduzidos, devem no entanto ser preservadas as características da linha. O espaço do mapa disponível para representar as entidades selecionadas é uma função da escala.



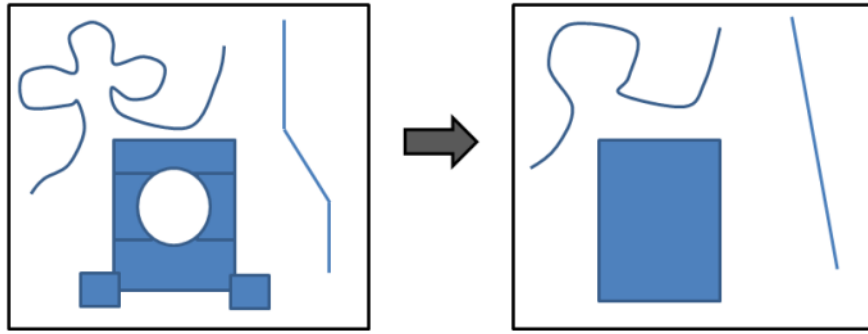


Figura 2.11 - O operador de simplificação.

Este processo visa assim, reduzir o detalhe de objetos espaciais. O processo da redução é o mais comum para objetos lineares e areais.

#### 2.4.6 Suavização

Em dados vectoriais as curvas dos objetos lineares que são demasiado angulosas podem ser alisadas ou suavizadas aplicando algoritmos de suavização, mantendo apenas as tendências mais significativas da linha. Tais processos adicionam nós ou vértices entre os pontos existentes de tal maneira que fazem a linha completa parecer mais lisa, ajudando a reduzir os ângulos acentuados. Normalmente a versão suavizada oferece uma representação estética mais agradável, tornando o mapa mais legível, ver figura 2.12.

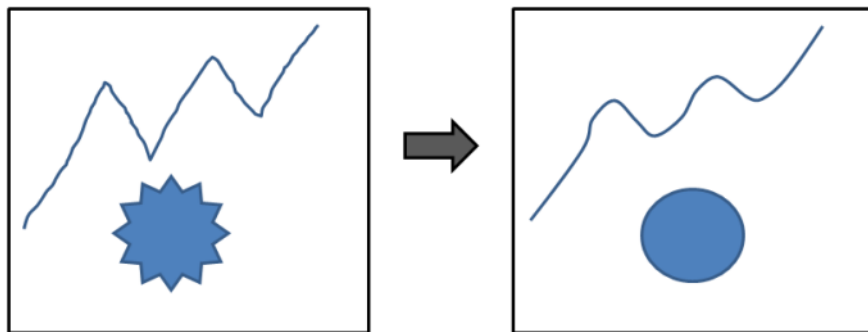


Fig 2.12 - O operador de suavização.

#### 2.4.7 Realce

As formas e tamanhos de entidades por vezes precisam ser exagerados ou enfatizados para satisfazer as exigências específicas de um mapa. Com os processos da generalização os objetos lineares podem tornar-se mais lisos, podendo haver necessidade de realçar partes de objetos que se pretenda evidenciar, como por exemplo curvas em estradas. A diferença para a

transformação exagero, cf. 2.4.12, é que o operador de realce funciona a maior parte das vezes, com a simbolização de entidades.

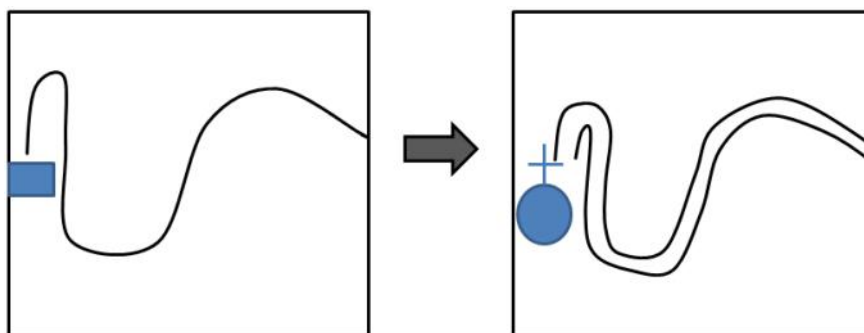


Figura 2.13 - O operador de realce.

### 2.4.8 Fractalização

As entidades tais como linhas de costa são normalmente muito complexas e a sua suavização, ou realce de detalhes, na escala maior podem ser conseguidos aplicando técnicas fractais, onde cada componente individual de uma linha suavizada na escala maior é uma réplica exata do componente da linha na escala menor. A definição fractal da linha força, ou obriga, normalmente os ângulos menores ao longo da linha a tornarem-se maiores efetuando a suavização da linha, figura 2.14.

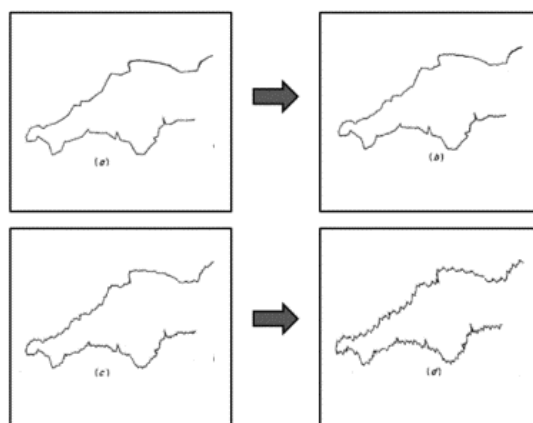


Figura 2.14 - O operador de fractalização (adaptado de Falconer, 1990).

### 2.4.9 Fusão

Quando a mudança de escala é significativa, por vezes é impossível preservar as características individuais de entidades lineares. Então, estas entidades lineares devem ser fundidas. Esta transformação é chamada fusão. De certo modo, a fusão é semelhante a

agregação pois ambas unem entidades, mas a agregação, cf. 2.4.3, efetua-se em entidades adimensionais, a fusão opera em dados unidimensionais; i.e. pontos versus entidades lineares.

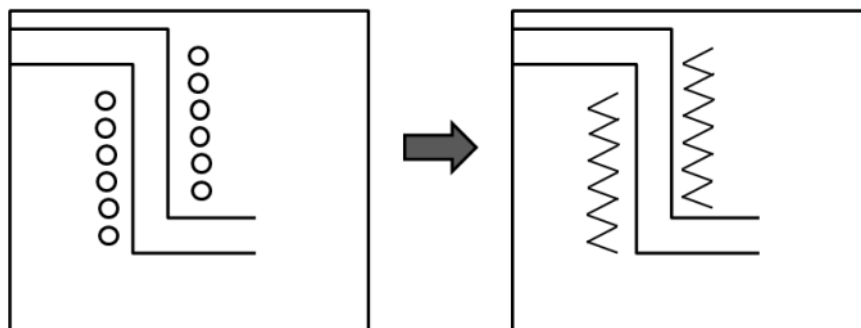


Figura 2.15 - O operador de fusão.

#### 2.4.10 Amalgamação

A transformação de amalgamação assemelha-se às transformações agregação e fusão, mas enquanto estas operam em entidades adimensionais e unidimensionais, a amalgamação opera com entidades bidimensionais, ou seja, áreas. Na amalgamação de áreas num elemento maior, é frequentemente possível reter as características gerais dessas áreas apesar de uma redução de escala. Por exemplo, se um mapa cobrir uma área que contém muitos lagos pequenos, esses lagos podem judiciosamente ser amalgamados em lagos maiores, e reter as características originais do mapa quando a escala é reduzida.

A figura 2.16 ilustra este processo de transformação onde três pedaços de solo a sul da estrada estão separados por um pequeno rio. Quando a escala é reduzida, estas áreas fundem-se. Para evitar este efeito, as três áreas de solo a sul da estrada (todos eles têm os mesmos atributos) são amalgamadas numa área grande de terra. Como mostra a figura, o resultado é um mapa menos complexo, mais fácil de entender. A supervisão humana deste processo é frequentemente necessária para evitar a amalgamação de objetos que devem reter as suas características espaciais individuais.

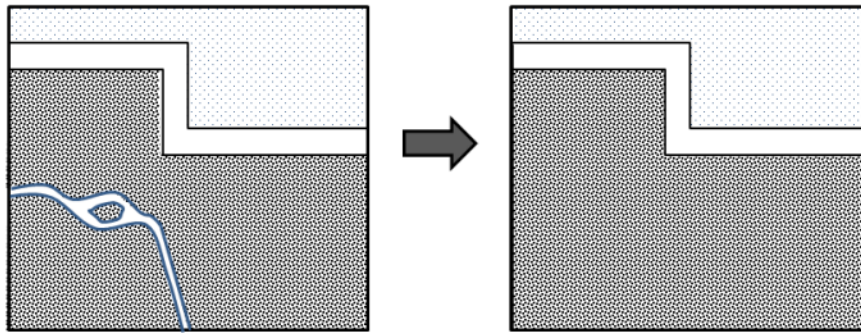


Figura 2.16 – O operador amalgamação

### 2.4.11 Colapso

A transformação de espaço chamada colapso, representa uma determinada entidade com uma redução de dimensão dessa entidade. Em mapas de pequena escala muitos objetos areais não podem ser representados proporcionalmente à sua escala. Em vez disso são representados por pontos ou linhas. Por exemplo na figura 2.17 onde um rio, alarga num lago. Quando a escala do mapa é reduzida, as margens do rio fundem-se. Como o rio, que é representado por uma entidade linear, alarga num lago que é representado por uma entidade areal, há um ponto de ramificação, onde os dois lados do rio podem ser delineados claramente sem coalescência do espaço entre as duas derivações da linha.

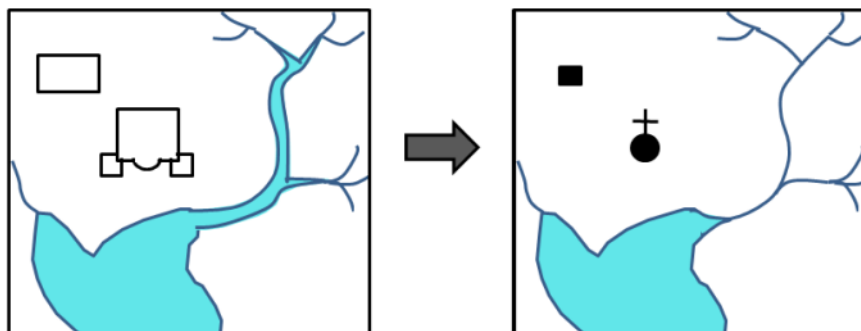


Figura 2.17 - O operador de colapso.

### 2.4.12 Exagero

A transformação espacial chamada exagero é usada para mudar as formas e tamanhos de entidades, de modo que são exageradas para satisfazer as exigências específicas de um mapa. Isto é frequentemente motivado pelo facto do tamanho físico de muitos objetos do mapa, não

permitir que sejam representados de forma conveniente. O exagero não é mais do que o realce gráfico de características significativas de entidades do mapa tais como a ampliação da largura da estrada, ver figura 2.18, ampliação de um edifício, ou parte dele, que seja de interesse arquitectónico, ou ainda caricatura de uma forma num sentido predefinido.

Em mapas de pequena escala muitos dos objetos que são importantes para serem representados, são pequenos na realidade, tal como estradas, edifícios ou pontes. Por esse motivo a sua representação terá que ser exagerada.

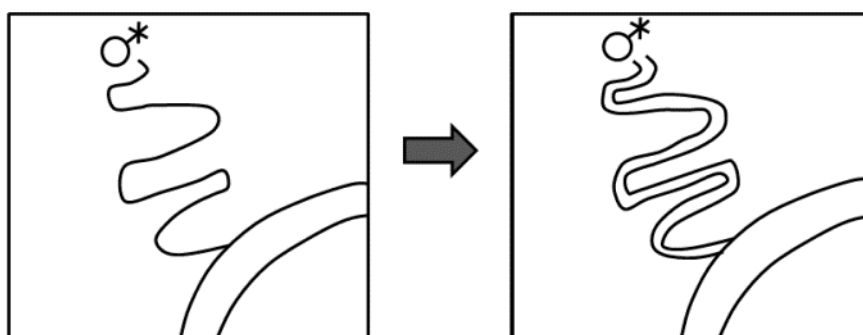


Figura 2.18 - O operador de exagero.

### **2.4.13 Tipificação**

Tipificação significa que um grande número de objetos discretos com formas similares são representados por um pequeno número de objetos que tem a mesma forma típica ver figura 2.19. Os objetos tipificados têm que preservar as características iniciais da distribuição. Aplica-se este operador quando a escala do mapa não permitir uma representação geométrica exata da entidade. Em alguns casos os pequenos objetos isolados de uma classe podem juntar-se noutra classe dominante.

Pretende-se assim reduzir a densidade das entidades e o nível de detalhe, mantendo a distribuição representativa do padrão e impressão visual do grupo da entidade original. Por exemplo, reduzir a importância do detalhe numa rede de drenagem sem perder a impressão visual desta estrutura.

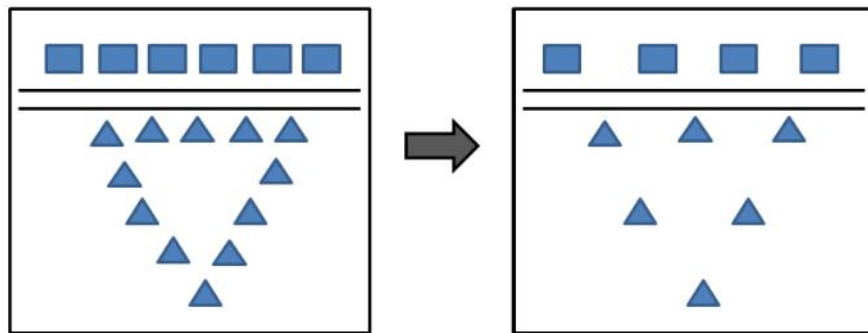


Figura 2.19 - O operador de tipificação.

#### **2.4.14 Deslocamento**

A transformação espacial chamada deslocamento é usada para evitar a fusão entre entidades, quando a colocação de duas ou mais entidades no mapa entram em conflito. A motivação aqui é ser capaz de usar simbolização de entidades mesmo quando a sua colocação física as sobrepõe ou fiquem muito aproximadas. Este processo é finalizado, geralmente de forma manual.

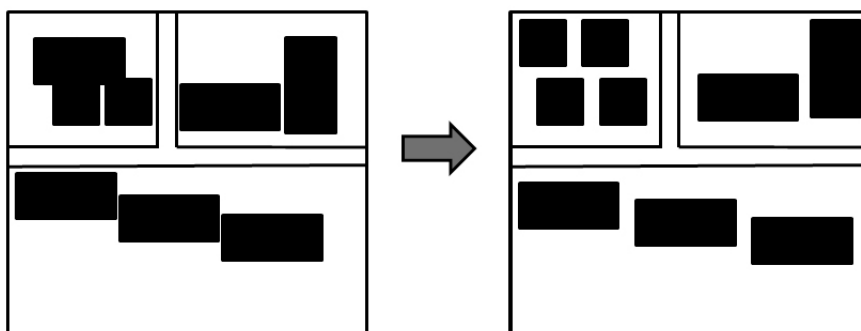


Figura 2.20 - O operador de deslocamento.

#### **2.4.15 Refinamento**

A transformação de refinamento seleciona um número e padrão de entidades que, ou são muito pequenas ou muitas para serem representadas claramente e, descreve-as duma forma que reduz a complexidade do mapa. Isto é normalmente feito omitindo as entidades menores, ou entidades que acrescentam pouco à percepção geral do mapa.

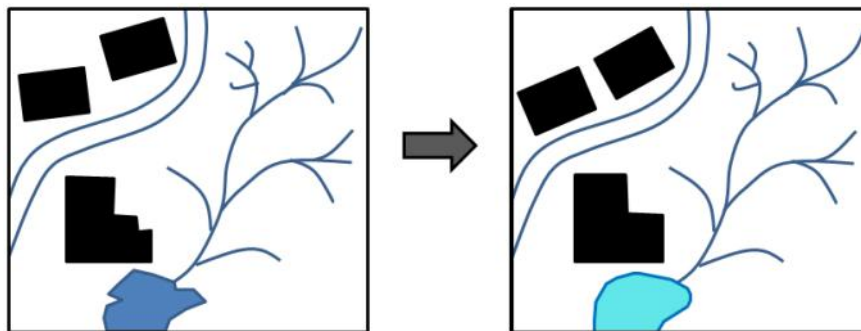


Figura 2.21 - O operador de refinamento.

### 2.4.16 Anamorfose

A anamorfose é uma transformação local de um conjunto de objetos a fim de resolver conflitos de proximidade. A anamorfose é composta de deslocamentos e deformações locais com propagação.

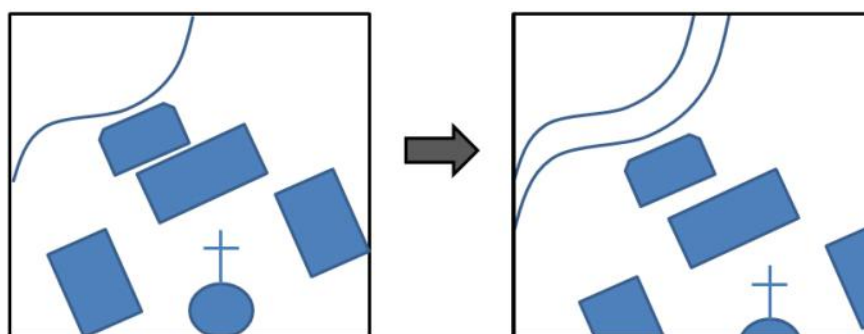


Figura 2.22 – O operador anamorfose

Ainda não existe um consenso na definição e no significado de operador (Rieger e Coulson, 1993), apesar da proposta de (Foerster et al. 2007) numa tentativa de formalização dos operadores de generalização. Embora faltem ainda alguns operadores como o realce ou estruturação/tipificação em software comercial, as plataformas atuais já permitem bons níveis de performance em generalização interativa (Lecordix et al. 1995). Alguns autores como (Lecordix et al. 1997) propuseram novos operadores como o balão e o acordeão, baseados em algoritmos com os mesmos nomes, para solucionar problemas de generalização de estradas sinuosas, nomeadamente em montanha, ver (Ruas, 2002).

Um operador define uma transformação que pode ser aplicada aos dados espaciais. A transformação depende do modelo de dados, e.g. raster ou vector, e depende da forma como é

feita a transformação, ou seja do próprio algoritmo. Existem algoritmos desenvolvidos para executar a simplificação (Reumann e Witkam, 1974; Lang, 1969; Douglas e Peucker, 1973; Cromley e Campbell, 1992), seleção (Töpfer e Pillewizer, 1966; Ruzak e Castner, 1990; Thompson e Richardson, 1995; Ruas 1999), e deslocamento (Lichtner, 1979; Mackaness, 1994; Roberts, 1997; Jones et al. 1995; Ruas, 1998; Burghardt e Meier, 1997; Nickerson, 1988; Bader e Weibel, 1997; Lonergan e Jones, 2001; Bader e Barrauld, 2001). Além disso, alguns algoritmos são genéricos, enquanto outros foram desenvolvidos para tipos específicos de objetos, tais como edifícios ou estradas (Cecconi, 2003). Consequentemente a escolha do algoritmo apropriado para cada operador é importante a fim de efetuar a melhor transformação num conjunto de dados espaciais. Esta escolha é geralmente feita de forma empírica. No entanto, pode ser baseada na eficiência e na exatidão esperadas do processo de generalização cartográfica (McMaster e Shea, 1992). A lista de algoritmos é enorme, mais ou menos complexos, com diversos parâmetros por algoritmo, que só com experiência se podem prever resultados.

## **2.5 Generalização de linhas**

As curvas são os elementos mais abundantes em cartografia. As linhas estão presentes na cartografia como elementos lineares e também como o limite de elementos areais. Assim a existência de técnicas de generalização adequadas para elementos lineares é essencial na generalização de um mapa ou BDG. Quando executamos tarefas de generalização, como simplificação ou suavização de linhas é importante que as relações espaciais entre entidades sejam mantidas. Por exemplo, entidades coincidentes devem permanecer coincidentes, o ponto de intersecção de duas entidades lineares e respetiva localização deve ser mantido, assim como as suas relações topológicas (Smith et al. 1999).

Do ponto de vista cartográfico uma linha não é uma abstração geométrica mas sim a representação de uma entidade geográfica representando uma curva de nível, uma estrada, uma linha de costa, ou o limite de um edifício ou outro objeto. Estas linhas devem ser tratadas de forma diferente, conforme as suas características intrínsecas. Algumas metodologias propostas por investigadores incluem métodos de segmentação, que dividem a entidade linear complexa, em partes homogéneas. Esta segmentação é um pré-requisito para a aplicação local de muitos algoritmos.



Como a seleção de operadores, algoritmos e respectivos parâmetros é crítica para o processo de automatização da generalização, vários autores têm utilizado uma abordagem algorítmica em vez de investigar modelos conceituais efetivos. Além disso, existem inúmeras investigações, tanto teóricas quanto práticas, sobre algoritmos para generalização de linhas. Müller (1986) investigou a influência da generalização cartográfica sobre a dimensão fractal de linhas que representam entidades geográficas. Como os cartógrafos tentam preservar o grau relativo de complexidade entre os vários segmentos de linhas de um mapa como uma medida para a generalização, este autor propôs o uso da geometria fractal para relacionar a complexidade de uma linha com o seu comprimento. As suas experiências indicaram que a dimensão fractal foi afetada pela mudança de escala em quase todas as linhas testadas. Lopes (2006) constatou que a dimensão fractal das curvas de nível diminui com a generalização, concordando com as conclusões apresentadas por (João, 1998) no seu estudo, em que *o comprimento das entidades lineares diminui com a crescente generalização*.

Os operadores mais utilizados na generalização de linhas são os operadores de simplificação e de suavização. O algoritmo de Douglas Peucker um dos mais utilizados para simplificação de linhas, sendo ainda frequente que resultados de novos algoritmos propostos sejam comparados com este. Não se pode afirmar que a maior redução no número de pontos represente um benefício real para a generalização de linhas. Contudo, faz sentido preservar os pontos extremos de cada linha original, o que ajuda a evitar problemas topológicos potenciais, no entanto não nos podemos esquecer que é necessário preservar a forma dos objetos e assim há um número de pontos que deveremos manter.

O processo de generalização, conforme observado na natureza, mostra que os objetos se tornam progressivamente menores à medida que se distanciam de um observador. Li e Openshaw (1992) descreveram um conjunto de algoritmos para o que eles chamaram de generalização de linhas localmente adaptativa. Considerando-se um mapa em qualquer escala, sempre haverá um tamanho mínimo de objeto cartográfico além do qual qualquer informação detalhada se perde, ou deixa de fazer sentido. Uma vez definido o tamanho mínimo visível ou observável, percorre-se a linha com o objetivo de eliminar pontos que ficam dentro da distância coberta pelo tamanho mínimo.

Outros autores, como Visvalingham e Whyatt (1993), utilizaram o conceito de área efetiva para realizar a simplificação progressiva de linhas através da eliminação de pontos. Neste

conceito o tamanho dos objetos, expresso por medidas de área, é a métrica mais fiável para a eliminação de pontos, já que as distâncias entre pontos e as medidas angulares são consideradas simultaneamente quando se usa o conceito de área. O método elimina iterativamente o ponto que produz o deslocamento de área mínimo relativamente à linha parcialmente simplificada.

Para a obtenção de curvas de nível para as escalas menores, alguns autores utilizaram uma abordagem diferente, baseada na generalização do MDT (*Modelo Digital do Terreno*) e obter as curvas de nível a partir do modelo generalizado (Li et al. 1999). A vantagem desta metodologia é que há a garantia de não haver intersecções entre curvas de nível, se forem aplicados algoritmos de interpolação de curvas de nível. No entanto os problemas associados a esta metodologia, são conceptualmente diferentes daqueles que o cartógrafo enfrenta na generalização de linhas. Por um lado temos a generalização do próprio MDT, que enfrenta problemas diferentes sendo modelo TIN (*Triangular Irregular Network*) ou GRID (*Grelha Regular de pontos*), há ainda a considerar a vetorização e extração das curvas de nível a partir do MDT generalizado.

### **2.5.1 Segmentação de linhas**

Na generalização automática de linhas, o desempenho de toda a operação de generalização depende do tipo e propriedades da linha. A descrição e classificação da linha é uma etapa essencial a fim de enriquecer as bases de dados e desta forma facilitar a tomada de decisão. Além disso, esta caracterização de linhas envolve uma tarefa de segmentação em secções homogéneas, porque cada linha pode ter uma geometria heterogénea e a descrição não seria aplicável a toda a linha. A importância da segmentação de linhas foi enfatizada por diversos autores incluindo Dutton (1999), Richardson e Mackaness (1999), Plazanet (1995, 1997), Woojin e Kiyun (2011), Nakos et al. (2008), Visvalingham e Williamson (1995), Balboa e López (2008, 2009). É evidente que seria útil ter uma segmentação automática, a fim de aplicar a cada secção o melhor algoritmo e os parâmetros apropriados. Consequentemente, a segmentação ótima, com um processo de enriquecimento da secção, é o primeiro requisito para adaptar processos, algoritmos e parâmetros aos diferentes tipos de linha, utilizações e escalas. Balboa e López (2009) apresentaram uma metodologia de segmentação baseada em reconhecimento da sinuosidade medida por meio da área eficaz, derivada do algoritmo de

Visvalingam Whyatt. As secções foram determinadas aplicando o algoritmo de Douglas Peucker.

A segmentação de linhas, segundo Balboa e López (2009) pode ser efetuada recorrendo a:

- Segmentação manual: Neste caso o operador executa a segmentação recorrendo aos seus critérios visuais e subjetivos, preferivelmente sobre o mapa em papel. As vantagens seriam as derivadas do processamento holístico e do conhecimento perito da mente humana. As desvantagens são as típicas de um processo humano: falta de repetitividade, subjetividade, análise qualitativa em vez de quantitativa, etc..
- Segmentação automática: Neste caso podiam ser usadas metodologias algorítmicas em computador. Tendo como vantagens ser repetitivo, objetividade, análise quantitativa, etc.. As desvantagens prendem-se basicamente com a falta da aplicação do conhecimento cartográfico.
- Segmentação semiautomática: Um perito controlaria a segmentação executada pelo algoritmo computacional, de modo a garantir a homogeneidade das secções. O sistema digital não é mais do que uma ferramenta auxiliar para o cartógrafo, este método tem a vantagem de combinar as duas opções precedentes.

O resultado da aplicação de um algoritmo a uma linha particular depende muito das características dessa linha (e.g. sinuosidade, complexidade, ...). Contudo, as características da linha não são constantes. Quando as linhas não são homogéneas, nenhum conjunto de valores de parâmetros utilizado no algoritmo de generalização assegura a generalização adequada de toda a linha (Byong-Nam e Young-Gul, 2007; Mackaness e Steven, 2006). A segmentação automática de linhas visa, num objetivo mais vasto, a automatização do processo da generalização. Alguns algoritmos e valores de parâmetros são mais apropriados para determinados tipos de linha, além disso uma linha não homogénea, deve ser segmentada em secções mais homogéneas para permitir melhorar os resultados da generalização.

Ruas (2002) descreve a caracterização e segmentação de estradas baseada num método hierárquico em testes de complexidade e homogeneidade das linhas. O método foi implementado na plataforma PlaGe (*Platform for research in Generalization*), (Plazanet et al. 1996). Este método começa por detetar os pontos característicos da linha para a sua segmentação e análise, tais como: descontinuidades na curvatura, pontos iniciais e finais, máximo de curvatura (vértices), mínimos de curvatura ou pontos de inflexão e pontos críticos

escolhidos entre os pontos de inflexão. De qualquer modo, as medições seguintes são consideradas para segmentação e análise, ver figura 2.23:

- A altura ( $h$ ) ou superfície do triângulo (definido pelos pontos de inflexão e o ponto mais afastado da linha assim definida);
- A distância euclideana entre os pontos de inflexão ( $base$ );
- O comprimento da curva entre pontos de inflexão ( $l$ );
- A razão entre o comprimento da curva e a distância euclideana entre os pontos de inflexão ( $l/base$ );
- O ângulo entre as tangentes nos pontos de inflexão;
- O número de vértices secundários por curva;
- A área destes triângulos e triângulos secundários.

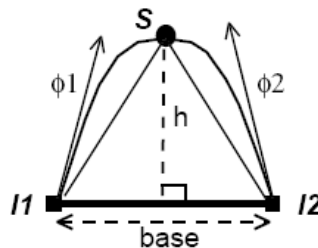


Figura 2.23 – Cálculo das medidas de caracterização utilizada por Plazanet (1996)

Informação adicional pode ser vista em Plazanet (1995), Plazanet et al. (1996), Battersby e Clarke (2003) e Cetinkaya et al. (2006).

## 2.5.2 Caracterização de linhas

Controlar a forma como a linha é alterada durante a generalização é da máxima importância, para manter a exatidão posicional e o reconhecimento, que são dois dos objetivos preliminares da generalização (Bernhardt, 1992). Pode dizer-se que linhas diferentes respondem diferentemente a um dado algoritmo de simplificação, e que do mesmo modo, uma dada linha responde também diferentemente aos diferentes algoritmos de generalização.

O processo de generalização é melhorado se a linha for caracterizada antes da sua entrada no sistema responsável pela generalização. O processo de generalização pode ser desta forma

adaptado à linha particular, usando as medidas quantitativas que caracterizam a linha. O objetivo é então selecionar a melhor combinação de algoritmos a ser aplicada e os valores apropriados para os parâmetros a utilizar nos respectivos algoritmos. Um grande número de parâmetros pode trazer grandes problemas para esta caracterização e consequente recolha de dados para este propósito. Para ser eficiente teremos que encontrar o número ótimo de medições das características.

McMaster (1986) definiu 7 categorias, para a caracterização: Comprimento da linha, valor das coordenadas, angularidade, curvilinearidade, vetor diferença, polígono diferença e perímetro de acordo com o tipo de informação que é analisada. Jasinski (1990) propôs o seguinte conjunto de parâmetros: média do comprimento dos segmentos e desvio padrão respectivo, média da distância entre a linha original e a generalizada e também o seu respectivo desvio padrão, média e desvio padrão da angularidade, curvilinearidade e dimensão fractal. Para caracterizar linhas Bernhardt (1992) usou os seguintes parâmetros caracterizadores: Plot de Richardson, dimensão fractal, angularidade, rácio de curvilinearidade e afastamento da linha base. Battenfield (1992) estudou o conjunto formado pelos seguintes parâmetros: Plot de Richardson, largura de banda (largura do retângulo exterior á linha, em que dois dos lados do retângulo são paralelos á linha base); segmentação, variância do erro (área entre a linha base e a curva); concorrência (número de vezes que a curva passa a linha base). Sendo a linha base referida como a linha reta que une o primeiro e o último ponto da curva.

É impossível caracterizar totalmente a forma de uma linha com uma simples quantidade. No entanto, há parâmetros que se adaptam melhor a determinados casos do que a outros, podendo mesmo alguns deles, em situações concretas, serem dispensados, por não trazerem qualquer vantagem para a caracterização ou posterior utilização. Além disso a quantidade de parâmetros para caracterizar completamente uma linha tornar-se-ia incomportável de modo que não seria prático, no sentido de os adquirir e utilizar em tempo útil. Um grande número de parâmetros poderia aumentar os custos de processamento inviabilizando a generalização, em determinados casos, como generalização *on-the-fly*.

### **2.5.3 Classificação de linhas**

A classificação das linhas, ou das partes em que as mesmas foram segmentadas, é um processo que tem em vista a aplicação dos mesmos algoritmos e respectivos parâmetros, às

linhas em questão, com a esperança de obter os mesmos resultados nas linhas da mesma classe. Esta é uma visão que não acontece para muitos casos, pois aqui dever-se-iam ter em conta as situações geográficas, que são sempre diferentes, muitas vezes fruto de objetos vizinhos, ou de outros que de alguma forma interagem com os objetos lineares em questão.

O sistema para a generalização automática deve analisar a forma e a geometria de cada linha num nível global, tal como num nível local (Nakos et al. 2008). Balboa et al. (2008), Balboa e López (2008) consideram a classificação de estradas divididas em cinco grupos, “*Muito suave, Suave, Sinuoso com direção estável, Sinuoso com direção variável e Muito sinuoso*”, ver (Vázquez, 2003; López e Balboa, 2007). Skopeliti e Tsoulos (1999) efetuaram dois tipos de classificação, classificação hierárquica e classificação não hierárquica, para classificar segmentos da linha de costa das ilhas de Ithaca e Lefkada. Classificaram as linhas em quatro grupos aos quais aplicaram diferentes algoritmos e parâmetros. As classes utilizadas foram: suave, muito suave, sinuoso e muito sinuoso. Na classificação hierárquica o número dos grupos indica a variedade das formas, que existem nos dados generalizados e na habilidade de os distinguir entre eles. Na classificação não-hierárquica o número dos grupos indica o grau de generalização. O número de grupos de classificação de linhas diminui enquanto a forma das linhas é simplificada.

Os processos de manipulação e armazenamento de dados nas BDG, a forma como estas são pesquisadas e ainda o paradigma associado à transformação algorítmica desses dados, inviabilizam a visão holística de todo o processo e da área a generalizar. Se a solução algorítmica não existir, todo o esforço de segmentação e classificação pode ter sido em vão. A segmentação da linha aumenta a pureza das classes utilizadas em IA, sendo muito útil para a sua classificação. A classificação pode ser efetuada automaticamente recorrendo a técnicas de IA, ver ponto 3.2.1 do Capítulo 3, sobre classificação supervisionada e classificação não-supervisionada.

## **2.6 Estratégias para a generalização automática**

Devido à grande complexidade da automatização da generalização, têm sido propostas algumas estratégias para abordar o problema. Segundo (Cecconi, 2003) as estratégias para a generalização cartográfica automática, dividem-se em:

**a) Estratégia orientada ao processo**

O processo da generalização resulta da derivação de dados novos, em tempo real, de uma única base de dados detalhada. Esta derivação é feita no que diz respeito à escala e ao tema do mapa (Weibel, 1997). Só existem soluções parciais com esta estratégia, devido à complexidade do processo da generalização (Cecconi, 2003).

**b) Estratégia orientada à representação**

Esta estratégia consiste no desenvolvimento de uma base de dados multi-escala que abrange mapas em escalas fixas diferentes. Evita-se desta forma a complexidade do processo cartográfico. Contudo, introduz problemas ligados à manutenção da base de dados e inconsistências entre as diferentes escalas.

**c) Estratégia orientada à derivação**

Este caso consiste no desenvolvimento de uma base de dados multi-escala ajustada, aplicando um processo de generalização externo numa única base de dados detalhada. Esta estratégia funde as duas estratégias acima. Similar à estratégia orientada à representação, sendo composta de níveis de detalhe diferentes, os quais, entretanto, são derivados de uma base de dados detalhada aplicando um processo de generalização. Assim, a série de dados é consistente e os objetos correspondentes a níveis diferentes são conectados, construindo uma estrutura hierárquica.

Estas estratégias possuem consequentemente, vantagens e desvantagens, que podem tornar o processo inflexível, no que diz respeito à escala e ao tema do mapa. Nenhuma das estratégias cumpre inteiramente todos os requisitos para o uso de generalização de mapas, em tempo real, na web, uma estratégia nova para criar mapas poderá vir a ser desenvolvida compreendendo as vantagens enunciadas.

Ruas e Plazanet (1996) propuseram uma estratégia para a automatização da generalização baseada no modelo de Brassel e Weibel, complementado com uma abordagem recursiva de *“iniciar com um esboço vago do que se pretende, aplicar um conjunto de operadores de generalização à fonte de dados, obter e analisar o resultado, repetir e refinar a aplicação subsequente de operadores de generalização num ciclo, até que seja encontrada uma solução satisfatória”* proposta por (Mackaness, 1995). O modelo de Brassel e Weibel especifica que os objetos necessitam de ser descritos corretamente, a fim de encontrar um método apropriado para os generalizar. Incluindo a descrição geométrica de um objeto tal como uma descrição

espacial de um conjunto de objetos. Ruas e Plazanet (1996) utilizaram na sua estratégia o conceito dos *confinamentos* para limitar o espaço do problema, de acordo com as especificações da base de dados ou do mapa.

Introduziram ainda o conceito de situação, que é descrita:

- 1) pelos objetos geográficos envolvidos;
- 2) pelos seus relacionamentos e;
- 3) pelas violações dos confinamentos.

A fim de usar uma situação como um inicializador para uma ação de generalização, esta tem que ser bem caracterizada. Contudo, mesmo quando as situações são bem descritas, o plano escolhido pode produzir resultados indesejáveis.

A estratégia para a generalização proposta por (Ruas e Plazanet, 1996), começa com a concepção de um plano global que representa a parte determinística do processo, isto é uma programação global das ações a executar, ver figura 2.24.



Fig 2.24- Plano geral da estratégia proposta por Ruas e Plazanet (adaptado de Ruas e Plazanet, 1996)



No nível mais baixo, uma transformação é o resultado de um algoritmo aplicado a um ou mais objetos. O objetivo é poder escolher o conjunto mais pertinente de objetos, algoritmos e parâmetros de acordo com um conjunto de critérios e do conhecimento adquirido, realizar a transformação e validá-la, por meio de algoritmos de controlo. Para efetuar estas escolhas, necessitamos de um plano global de ações, com o objetivo de saber o tipo de transformações que devem ser executadas e o tipo de objetos envolvidos em cada transformação. Este plano global não será suficiente para resolver todos os problemas. Embora as tarefas subsequentes sejam mais simples. É de salientar que a solução algorítmica poderá não existir, desta forma apenas a generalização manual efetuada por um cartógrafo humano poderá solucionar o problema, informação adicional pode ser vista em Brenner e Sester (2005).



# Capítulo 3

## Conhecimento e Inteligência Artificial

---

### 3.1 Introdução

O processo de decisão envolvido na generalização cartográfica é extremamente complexo. Como descrever corretamente e representar a informação geográfica e como converter o conhecimento e a experiência dos cartógrafos em conhecimento que os computadores possam adquirir, utilizar e manipular são as chaves de uma generalização cartográfica inteligente. Neste capítulo, debruçamo-nos sobre a temática do conhecimento utilizado no contexto da generalização cartográfica e da Inteligência Artificial. Posteriormente abordam-se as técnicas de aquisição e representação do conhecimento e finalmente apresentam-se os métodos de IA.

### 3.2 Conhecimento em generalização cartográfica

A abordagem mais frequente para automatização da generalização consiste na utilização de um método local, sequencial baseado no conhecimento (Brassel e Weibel, 1988). Nesta abordagem, cada objeto vetorial, caracterizado por uma série de coordenadas, que definem a sua geometria, é transformado aplicando uma sequência de algoritmos de generalização, que executam transformações geométricas atômicas. A sequência algorítmica não é predefinida mas vai sendo construída para cada objeto, controlada pelo conhecimento (heurística), dependendo das suas características e dos efeitos previstos, causados nos dados, pela utilização dos algoritmos. Esta abordagem controla uma base de conhecimento, em particular, adapta o conhecimento quando há novos elementos, tais como novos algoritmos de

generalização, que são integrados no sistema de generalização ou quando mudam as exigências do utilizador ou seja as especificações do mapa.

A investigação em generalização cartográfica tem mostrado que os métodos algorítmicos são adequados para pequenas tarefas, mas parecem ter um potencial limitado para resolver o processo global da generalização. As tentativas de usar sistemas baseados na representação do conhecimento explícito, por exemplo, sistemas baseados em regras ou sistemas periciais, têm tido igualmente um sucesso limitado (Meng, 2003; Weibel et al, 1995; Mustière et al, 1999; López e Balboa 2007). O principal fator de limitação dos sistemas de conhecimento explícito na generalização é a escassez de conhecimento formalizado disponível. Isto é, a aquisição de conhecimento, torna-se na principal dificuldade ao progresso de técnicas baseadas em conhecimento. De acordo com Meng (2003) a formalização inadequada do conhecimento foi identificada como um problema na implementação da generalização em BDG.

Há a considerar diversas fases na aquisição do conhecimento tal como a identificação do conhecimento, a extração das respetivas fontes, a codificação em formulário simbólico, organização e modificação para obter o melhor desempenho (Marshall, 1990). A aquisição do conhecimento consiste na extração do conhecimento especialista a ser transposto para um sistema computacional de modo a torná-lo inteligente. No caso da generalização cartográfica, envolve a pesquisa e a interpretação de processos manuais de generalização. As fontes para a aquisição deste conhecimento são variadas desde as provenientes de especialistas, bibliografias de confiança ou mapas generalizados manualmente. O grande desafio é traduzir o pensamento do cartógrafo num conjunto de processos explícitos e bem definidos para a implementação computacional. Também os exemplos da engenharia reversa, podem ser utilizados para adquirir o conhecimento da generalização através da comparação de objetos do mapa em diversas escalas (Buttenfield e McMaster, 1991; Leitner e Buttenfield, 1995; Weibel, 1995). Outros estudos geraram regras de generalização interativa realizada por um perito cartógrafo (Weibel, 1991; Weibel et al. 1995; McMaster, 1995; Reichenbacher, 1995). Diversos estudos foram efetuados na aplicação de técnicas de aprendizagem máquina para converter o conhecimento em especificações (Weibel et al. 1995; Plazanet et al. 1998; Mustière e Duchêne, 2001; Mustière, 2005; e Hubert e Ruas, 2003).

Müller e Mouwes (1990) estudaram uma série de mapas para iniciar uma introspeção em desafios para a generalização automática. Segmentaram o conhecimento em dois níveis: *conhecimento superficial* e *conhecimento profundo*. O conhecimento superficial é descrito

nas especificações do mapa, com significado para a generalização interativa, este conhecimento refere-se ao conhecimento transmissível em linguagem formal e sistemática. O conhecimento profundo<sup>3</sup> é mais importante, muito mais complexo de automatizar e é usado por cartógrafos, quando o conhecimento superficial não chega para a atividade efetuada, ou para resolução de problemas mais complexos.

### **3.2.1 Técnicas de aquisição de conhecimento**

#### ***a) Perícia humana e análise de documentos***

A forma mais simples de obter conhecimento na generalização cartográfica consiste na entrevista ao cartógrafo (Plazanet et al. 1998) na qual este é inquirido sobre a forma como lida com as várias situações de generalização. Não obstante, o conhecimento resultante por vezes pode não ser diretamente explorado para a automatização, pois necessita de uma interpretação da linguagem perita, que possa estar num nível conceptual relativamente elevado. O resultado é uma descrição parcial do processo em regras simples, sendo insuficiente na generalidade dos casos.

Uma comparação direta entre o mapa original e o mapa generalizado é um método interessante, desde que seja possível avaliar o tipo e o grau das modificações que se aplicam aos objetos nas diferentes escalas, necessitando no entanto de ser explicitado (Buttenfield, 1992), ou seja, explicado e colocado numa forma que possa ser utilizado. Este método, chamado de engenharia reversa, tenta identificar as operações que foram efetuadas para construir o mapa final. Contudo, considerando que geralmente o mapa final é o resultado de uma sequência de operações complexas, é difícil identificar inequivocamente a sequência apropriada das operações e muitas vezes até as próprias operações.

O registo de ações interativas do utilizador em sistemas automáticos, isto é, seguimento do processo de generalização, pode evitar problemas relacionados com a aquisição direta de conhecimento e análise de mapas e textos. Nestes sistemas, os operadores cartógrafos são confrontados com situações de generalização que resolvem interativamente, escolhendo os algoritmos apropriados e ajustando os seus parâmetros. As escolhas dos cartógrafos são armazenadas pelo sistema e podem ser usadas de alguma forma pelo algoritmo de

---

<sup>3</sup> **O conhecimento profundo** - compreende elementos cognitivos e modelos técnicos. Os elementos cognitivos são os modelos mentais, esquemas, paradigmas, crenças e pontos de vista que ajudam o indivíduo a perceber e definir o seu mundo, criando imagens da realidade e visões do futuro. Os modelos técnicos que o ser humano possui referem-se ao *Know-how*, técnicas e respetivas habilidades (Pereira, 2011).

aprendizagem. As vantagens deste tipo de método de aquisição vêm do facto que os cartógrafos expressam as suas escolhas diretamente de algoritmos da generalização e respetivos parâmetros, com um dispositivo automático de entrada, reduzindo a complexidade da transição do conhecimento perito ao software (Kazemi et al. 2005). No entanto, como as situações geográficas nunca são as mesmas e a situação concreta não é completamente descrita, é necessário a interpretação humana em cada caso, limitando os casos de sucesso.

### ***b) Aprendizagem máquina***

De acordo com Simon (1983) citado em Plazanet et al. (1998), com a aprendizagem denotam-se as mudanças no sistema que são adaptáveis, no sentido que permitem ao sistema executar a mesma tarefa ou tarefas, extraídas da mesma população, mais eficiente e eficazmente nas próximas vezes. As técnicas de aprendizagem dividem-se em duas estratégias principais: a aprendizagem supervisionada e aprendizagem não supervisionada.

#### ***b1) Aprendizagem Supervisionada***

Aprendizagem supervisionada é o processo onde as classes<sup>4</sup>, são previamente definidas e fornecidas ao sistema cognitivo assim como os exemplos de cada classe (Holsheimer e Siebes, 1994). O sistema tem desta forma que descobrir propriedades comuns dos exemplos ou padrões para cada classe, ou seja a *descrição da classe*. O objetivo do sistema é produzir um classificador que possa ser usado para prever a que classe pertencem os objetos. Mais formalmente, na aprendizagem supervisionada, é fornecido um conjunto de pares  $(x_i; y_i)$ , onde  $x_i$  é uma descrição e  $y_i$  é uma classe. O programa aprende ou determina uma função  $f$ , tal que  $f(x_i) = y_i$  para todo  $i$ . Ou seja, o conjunto de pares  $(x_i; y_i)$  designa-se conjunto de treino. O conjunto de treino é um conjunto de informação representado geralmente num formalismo do valor dos atributos, em que os dados são representados por um vetor de um conjunto fixo de atributos e etiquetados com um número de classes discretas.

A entrada do algoritmo de aprendizagem supervisionada é um conjunto de treino e a saída é um classificador. Este classificador pode ter em conta os diversos métodos da representação do conhecimento, por exemplo classificadores paramétricos e não paramétricos.

---

<sup>4</sup> **Classe:** Uma classe é a descrição de um conjunto de objetos que partilham os mesmos atributos, operações, relações e a mesma semântica. Uma classe corresponde a algo tangível ou a uma abstração conceptual (Silva e Videira 2001).

### ***b2) Aprendizagem não Supervisionada***

No caso da aprendizagem não supervisionada, as técnicas de aprendizagem, chamadas também aprendizagem de observação ou descoberta, o sistema tem que descobrir as classes, baseando-se em propriedades comuns dos objetos. Examina os exemplos de teste e verifica se exibem as mesmas regularidades, descobertas nos exemplos de treino. O resultado da aprendizagem não supervisionada, é um conjunto de descrições das classes, um para cada classe, que no seu conjunto cobrem todos os objetos. A aprendizagem não supervisionada é subdividida em dois grandes grupos: métodos hierárquicos e métodos de segmentação (Costa e Simões, 2004). Independentemente do algoritmo utilizado, teremos de ser capazes de julgar a qualidade dos agrupamentos obtidos. O critério geral é o de que deve haver a máxima semelhança entre os elementos do mesmo agrupamento e a máxima diferença entre elementos de agrupamentos diferentes. Implica que a noção de semelhança tenha de ser precisada.

É necessário enfatizar a importância de integrar o conhecimento dos cartógrafos com os sistemas de generalização. Isto facilitará o desenvolvimento de um sistema pericial, flexível e robusto, que seja capaz de gerar um método inovador para a generalização. Sem dúvida, uma avaliação detalhada dos sistemas de generalização e do seu desempenho são essenciais para englobar o conhecimento cartográfico dos peritos e trazer a inteligência artificial para uma estrutura de generalização.

## **3.3 Representação do conhecimento**

A comunidade da IA apresentou nos últimos cinquenta anos uma grande variedade de linguagens cada uma adaptada mais ou menos para representar um campo diferente do conhecimento dos seres humanos, que necessita ser representado e processado. Embora uma grande quantidade de perícia humana possa ser formulada como um conjunto de procedimentos ou de inferências específicas numa dada linguagem ou paradigma, o processo da generalização cartográfica exige claramente diversas linguagens de representação de conhecimento para capturar os diferentes tipos de conhecimento manipulados, variando desde os dados reais do mundo até à sua representação final como mapa útil.

O conhecimento é um pilar dos sistemas de IA. A representação do conhecimento é um componente fundamental em sistemas inteligentes. Através dos mecanismos de representação, formalismos da IA, o conhecimento é codificado através de objetos, atributos,

objetivos e ações, sendo processado através de estruturas e procedimentos (Bench-Capon, 1990).

Existem muitas formas de representação do conhecimento através de formalismos computacionais, como *scripts*, *frames*, *redes semânticas*, *regras*, *grafos conceptuais*, *representações formulário* e *conceitos, objetos e factos*. Estes formalismos podem ser utilizados para representar casos em sistemas de RBC (*Raciocínio Baseado em Casos*) e representar os domínios a utilizar nas regras (Weber et al., 1998). A seleção dos formalismos de IA depende da natureza do raciocínio e do próprio conhecimento. Para conferir inteligência à máquina é preciso conhecer os processos cognitivos da mente humana, com a finalidade de os modelar. Uma maneira de iniciar os estudos é através da compreensão dos processos do raciocínio humano. O raciocínio compreende um encadeamento aparentemente lógico de juízos ou pensamentos. É algo tão comum e intuitivo que a maioria das pessoas não se preocupa em analisar como tal processo ocorre (Rabuske, 1995).

Muitas das tarefas executadas nos processos cartográficos, especialmente em generalização cartográfica, são executadas intuitivamente por cartógrafos. Consequentemente, devido a uma falta de formalismo a automatização da generalização cartográfica, faz com que avanços concretos sobre generalização dependam de discussões sobre a representação do conhecimento cartográfico e sobre as possibilidades da sua implementação em sistemas periciais utilizados em generalização. Portanto, uma alternativa ao desenvolvimento e aplicação de algoritmos, é a tentativa de formalização e representação do conhecimento cartográfico através de modelos abrangentes e de técnicas de aquisição de conhecimento.

Mustière (2005), ao trabalhar com a generalização cartográfica de mapas de estradas, concluiu que o uso de diferentes algoritmos para estradas diferentes, ou para troços diferentes da mesma estrada, precisaria ser executado por regras claras, que definissem a utilização dos algoritmos durante o processo. O ponto crítico da sua abordagem é que o conjunto de regras depende da aquisição e da formalização do conhecimento cartográfico.

A chamada hipótese da representação do conhecimento serve de base aos trabalhos em IA. De acordo com essa hipótese, qualquer sistema, seja ele humano ou artificial, que demonstre um comportamento inteligente contém uma subestrutura básica de conhecimento (D'Algé, 2007). Essa subestrutura básica de conhecimento é manipulada por uma segunda subestrutura conhecida como motor de inferência, que processa os símbolos da subestrutura básica de



conhecimento com o propósito de gerar comportamento inteligente, conforme se pode ver na figura 3.1 (Kazemi et al. 2005), ver ainda (Reichgelt, 1991).

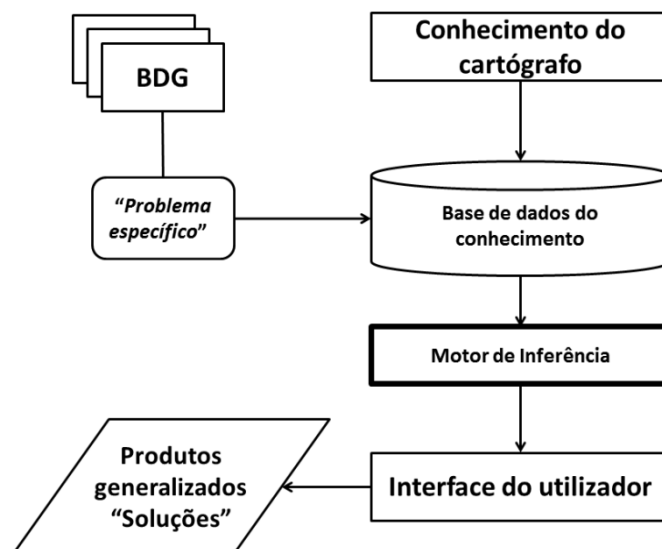


Figura 3.1 – Módulos de um sistema baseado em conhecimento (adaptado de Kazemi et al. 2005)

Presentemente, parece que é mais realístico caminhar na direção de sistemas de conhecimento cooperativos, que possam examinar o problema da generalização de forma holística, como uma tarefa partilhada entre especialista e máquina. As situações são sempre diferentes assim como os objetivos da sua representação, desta forma torna-se impossível a sua completa caracterização de modo a poder equacionar corretamente o problema e propor uma solução coerente.

Rieger e Coulson (1993) usaram um programa interativo de explicitação de conhecimento e entrevistaram cartógrafos com larga experiência profissional. Os resultados destas entrevistas indicaram uma falta de padrão no conhecimento declarado tanto no que se refere às definições de procedimentos para generalização, assim como no que se refere aos aspetos conceptuais. Assumindo que os sistemas periciais se devem basear no conhecimento consensual dos cartógrafos, este estudo indica que a área da generalização cartográfica não parece ser muito propícia a esta tecnologia.

Para D'Algé (2007) a discussão sobre representação do conhecimento, generalização cartográfica e sistemas periciais ilustra a importância da interdisciplinaridade das várias áreas envolvidas. A investigação futura em generalização deve incluir tentativas de entendimento e

esclarecimento do conhecimento geográfico, talento artístico e estético, senso de proporção, de harmonia, de formas e cores, estes são elementos indispensáveis para a criação de um mapa legível, claro e expressivo (Szego, 2011), ver ainda sobre este assunto (Bertin, 1983, 1999; Cartwright, 2011).

### 3.4 Inteligência Artificial

A Inteligência Artificial (IA) é um ramo da Ciência da Computação, dedicado ao estudo das técnicas que possibilitam a representação em máquinas, de algum aspeto da cognição humana (Weber, 1998). Segundo Rich (1988) *"IA é o estudo de como fazer com que a máquina possa realizar tarefas que hoje são melhor realizadas pelo homem"*. A Inteligência Artificial é desta forma vista como o ramo da ciência da Computação que investiga a criação de sistemas inteligentes. Existem duas abordagens em IA: uma científica, voltada ao estudo da psicologia cognitiva, cujo objetivo é compreender os processos envolvidos na inteligência e outra tecnológica, que estuda a representação destes processos através da máquina. Pelo facto da generalização cartográfica ser uma tarefa essencialmente cognitiva, têm surgido nos últimos anos propostas de adopção de inteligência artificial na resolução de problemas pontuais, baseando o conhecimento na aprendizagem de casos previamente conhecidos. As técnicas de IA utilizadas nesta tese visaram a predição do parâmetro a utilizar no algoritmo de generalização de curvas de nível, ver Capítulo 5, os seus *outputs* foram posteriormente analisados, cujo resultado prevê uma aplicação mais vasta a outros temas e cadeias de algoritmos.

### 3.5 Redes Neurais

A abordagem das *redes neuronais artificiais* consiste em capturar os princípios básicos de manipulação de informação do cérebro humano e aplicar esse conhecimento na resolução de problemas que exigem aprendizagem a partir da experiência (Beale et al. 1990).

O cérebro humano é um modelo que ocorre naturalmente quando pensamos na construção de máquinas inteligentes, ou com comportamentos inteligentes. Os estudos sobre redes neuronais artificiais são fundamentados sobre a estrutura do cérebro humano de modo a tentar emular a sua forma inteligente de processar informações, apesar de se desconhecer a maneira pela qual o cérebro manipula informações complexas (Beale et al. 1990). Tarefas

simples para o ser humano, são evidenciadas como complexas pela dificuldade encontrada no ensino de robôs a executar tais tarefas, e.g. como pegar num objeto, ou mesmo andar. A modelação do conhecimento utilizado na resolução de um problema específico, e.g. generalização de curvas de nível, pode ser representada através das conexões entre células nervosas. Estruturalmente, a rede neuronal artificial, também conhecida por modelo conexionista de computação, assemelha-se à rede neuronal biológica pela composição dos seus neurónios e pela sua interconexão (Wasserman, 1989).

De acordo com Gama et al. (2012), o primeiro modelo formal de redes neuronais artificiais foi proposto por McCulloch e Pitts, em 1943, conforme figura 3.2. Essa estrutura abstraía a complexidade da atividade neuronal em sistemas neuronais reais, assim como complicadas características encontradas no corpo de neurónios biológicos. Apesar destas redes iniciais não disporem de capacidade de aprendizagem, foram a base para a maioria dos modelos conexionistas desenvolvidos posteriormente.

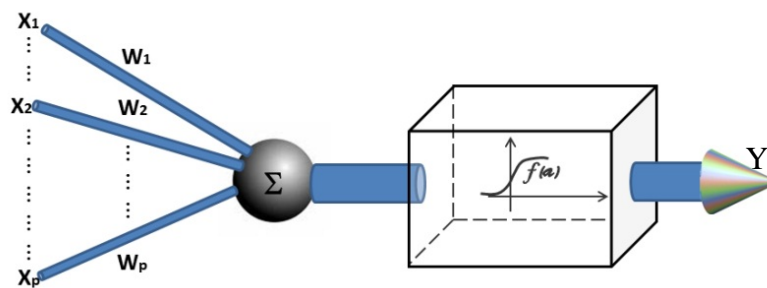


Figura 3.2 – Esquema de unidade de processamento, McCulloch-Pitts

Uma Rede Neuronal é constituída por um conjunto de unidades básicas de processamento designadas de neurónios ou nodos, conhecidos como unidades lógicas com limiar ou LTU (*Logic Threshold Unit*). Estes neurónios encontram-se ligados por conexões ou sinapses, em que o peso da conexão determina a intensidade desta ligação e é expresso por um valor numérico (Larose, 2005). O conhecimento, utilizado posteriormente em novos casos, é adquirido a partir dos dados num processo de aprendizagem e é armazenado nas conexões. O processo de treino, ou aprendizagem, é concretizado computacionalmente através de um algoritmo, na presente tese utilizou-se o algoritmo Backpropagation, um melhoramento do algoritmo Delta. Após o treino, a Rede Neuronal é capaz de responder, num curto intervalo de tempo, a novas situações (Costa e Simões, 2004). Uma iteração do algoritmo de treino é composta por ajustamentos dos pesos das conexões, para todos os casos de treino (Russel e Norvig, 2003). A operação de uma unidade de processamento, conforme figura 3.2, proposta por McCulloch e Pitts em 1943, pode ser resumida da seguinte forma:

- 1 - Os sinais são apresentados à entrada ( $x_1, x_2, \dots, x_p$ );
- 2 - Cada sinal é multiplicado por um número, ou peso, que indica a sua influência na saída da unidade, ( $w_1, w_2, \dots, w_p$ );
- 3 - É feita a soma ponderada dos sinais que produzem um nível de atividade;
- 4 - Se este nível de atividade exceder um certo limite (*threshold*) a unidade produz uma determinada resposta de saída, ( $y$ ).

### 3.5.1 Topologia de Redes Neurais

As redes neurais artificiais são sistemas computacionais distribuídos, compostos de unidades de processamento simples, densamente interconectadas. A topologia diz respeito ao modo como os diferentes neurónios se ligam entre si. Arquiteturas de redes neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

A estruturação de uma rede neuronal em camadas é uma importante característica topológica desses modelos (Másson e Wang, 1990). Numa rede neuronal estruturada em camadas, o conjunto de vértices  $V$  pode ser particionado em vários subconjuntos disjuntos  $V = V^{(0)} V^{(1)} \dots V^{(L)}$  de modo que as unidades de processamento de uma camada apresentem somente conexões com as unidades das camadas posterior e anterior. Quanto ao número de camadas, as redes neurais podem ser dispostas numa única camada, configuração mais simples de uma rede neuronal, ou em múltiplas camadas. Quando duas ou mais camadas são utilizadas, um neurónio pode receber nos seus terminais de entrada, valores de saída de neurónios da camada anterior e enviar o seu valor de saída para terminais de entrada de neurónios da camada seguinte. Os neurónios de uma camada podem ser conectados a todos os neurónios da camada seguinte ou apenas a parte deles, ver figura 3.3.

As camadas são classificadas em três grupos:

- Camada de Entrada: onde os padrões ou, neste caso, os vetores de atributos classificadores das linhas, são apresentados à rede;
- Camadas Intermédias ou Escondidas: onde é feita a maior parte do processamento, através das conexões ponderadas;
- Camada de Saída: onde o resultado final é concluído e apresentado.

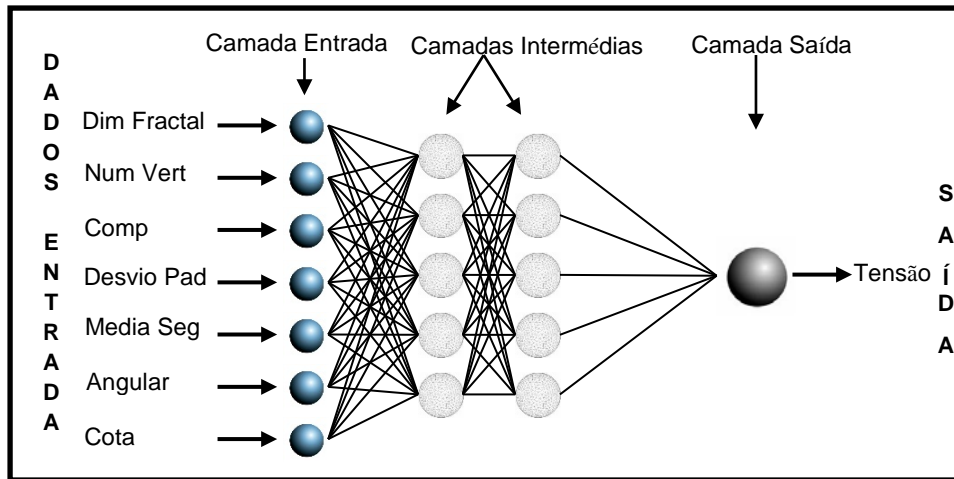


Figura 3.3 – Rede neuronal organizada em camadas. Neste caso tem uma camada de entrada e duas intermédias.

### 3.5.2 Aprendizagem em redes neurais

A maioria dos modelos de redes neurais possui uma determinada regra de treino, onde os pesos das suas conexões são ajustados de acordo com os padrões apresentados à rede. Por outras palavras, estas redes aprendem através dos exemplos ou padrões, que lhe são fornecidos. Os sinais podem ser positivos, chamados “*excitadores*” ou negativos, chamados “*inibidores*”. Uma entrada positiva promove a ativação do nodo, enquanto a negativa tende a manter o nodo inerte. Desta forma, os neurónios artificiais, ligados por conexões, trocam sinais inibitórios ou excitatórios, competindo ou cooperando entre si. O comportamento inteligente emerge então, da ação simultânea dessa coletividade, sem a necessidade de elementos centralizadores (Másson e Wang, 1990).

Uma das propriedades mais importantes das redes neurais é a habilidade de aprender e desse modo melhorar o seu desempenho. Isso é feito através de um processo iterativo de ajustamentos aplicado aos seus pesos, durante o treino. Existem muitos tipos de algoritmos de aprendizagem específicos para determinados modelos de redes neurais.

Na aprendizagem supervisionada é apresentado à rede, um conjunto de padrões de entrada e os seus correspondentes padrões de saída. Nos sucessivos processos, a rede efetua um ajustamento dos pesos das conexões, entre os elementos de processamento, segundo uma regra de aprendizagem (algoritmo), até que o erro entre os padrões de entrada e correspondente saída, esteja abaixo de um valor mínimo desejado, ver fluxograma da Figura 3.4. Os algoritmos de aprendizagem supervisionada mais utilizados são *Perceptron*, *Backpropagation* e *Adeline e Madaline* (Hetch-Nielsen, 1990). A aprendizagem por reforço é

similar à aprendizagem supervisionada com a diferença que a resposta fornecida pela rede é avaliada externamente (Braga et al., 2007).

Na aprendizagem não-supervisionada, a rede analisa os conjuntos de entrada apresentados, determina algumas das propriedades dos conjuntos de dados e aprende a refletir sobre as suas propriedades de saída. Os métodos de aprendizagem mais utilizados são: *Mapa Auto-Organizável de Kohonen*, *Redes de Hopfield* e *Memória Associativa Bidirecional* (Hetch-Nielsen, 1990).

O primeiro passo do processo de desenvolvimento de redes neuronais artificiais é a recolha de dados relativos ao problema e a sua separação num conjunto de treino e um conjunto de teste. Os dados de treino serão utilizados para o treino da rede e dados de teste serão utilizados para verificar a sua performance sob condições reais de utilização.

O segundo passo é a definição da configuração da rede, que pode ser dividido em três etapas, segundo (Braga et al., 2007):

- Seleção do paradigma neuronal apropriado à aplicação;
- Determinação da topologia da rede a ser utilizada (número de camadas, número de unidades em cada camada, etc);
- Determinação de parâmetros do algoritmo de treino e funções de ativação.

O terceiro passo é o treino da rede. Nesta fase, seguindo o algoritmo de treino escolhido, serão ajustados os pesos das conexões. Normalmente, os valores iniciais dos pesos da rede são números aleatórios uniformemente distribuídos num intervalo definido.

O quarto passo é o teste da rede. Durante esta fase o conjunto de teste é utilizado para determinar a performance da rede com dados que não foram previamente utilizados. A performance da rede é uma boa indicação de sua performance real (Braga et al., 2007). Finalmente, com a rede treinada e avaliada, pode então ser integrada no sistema em ambiente operacional da aplicação. Quando um padrão é inicialmente apresentado à rede, esta produz uma saída. Após medir a distância entre a resposta atual e a desejada, de acordo com uma métrica predefinida, efetuam-se os ajustamentos necessários nos pesos das conexões de modo a reduzir esta distância. Este procedimento é conhecido como Regra Delta ou regra de Widrow-Hoff, que pode ser apresentada da forma indicada na figura 3.4. Onde o par de treino (X, d) corresponde ao padrão de entrada e respetiva resposta desejada; O erro E é definido como, Resposta Desejada - Resposta Obtida (d - O);

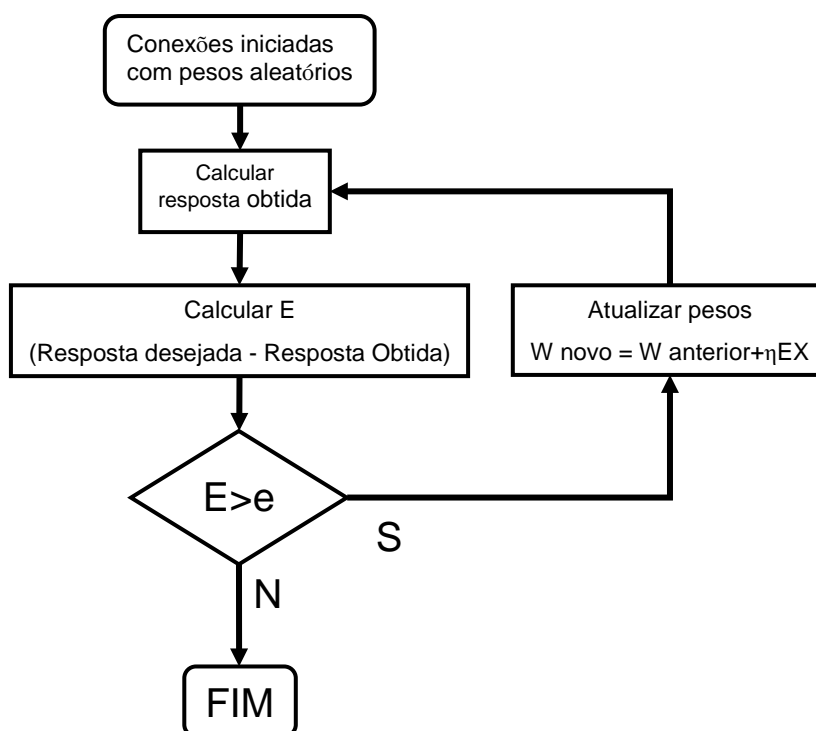


Figura 3.4 – Algoritmo de treino, Regra Delta

A taxa de aprendizagem  $\eta$  é uma constante positiva, ver figura 3.4, corresponde à velocidade de aprendizagem e controla a velocidade de convergência dos valores dos pesos para a situação desejada. Se  $\eta$  tiver um valor baixo, a convergência é lenta, se for elevado pode provocar oscilações. As respostas geradas pelas unidades são calculadas através de uma função de ativação. Existem vários tipos de funções de ativação, como representado na figura 3.5, as mais comuns são: *Hard Limiter*, *Threshold Logic* e *Sigmoid* ou *função logística*.

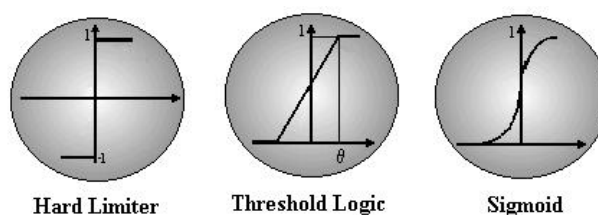


Figura 3.5 – Exemplos de funções de ativação

A função de ativação permite calcular o valor de saída do neurónio, a partir do valor de ativação. Segundo Nelson e Illingworth (1991) a função de ativação é geralmente não-linear. As funções lineares na prática mostram-se pouco eficientes pois fornecem simplesmente saídas proporcionais às entradas. Sendo as sigmóides as mais utilizadas.

De acordo com Beale et al. (1990), o *backpropagation* pode ser visto como uma generalização do método Delta, utilizado na rede Adaline (*Adaptive Linear*), para redes neurais de múltiplas camadas. Quando se apresenta um determinado padrão de entrada a uma rede neuronal não treinada e o respetivo padrão de saída, é produzida uma saída aleatória. A partir dessa saída produzida pela rede é calculado um erro, representando a diferença entre o valor obtido e o desejado. O objetivo consiste, então, em reduzir continuamente o erro até um determinado valor aceitável. Isto é alcançado pelo ajustamento dos pesos entre as conexões dos neurónios pela aplicação da regra Delta Generalizada, que calcula o erro para alguma unidade particular e propaga esse erro para a camada anterior. Cada unidade tem os seus pesos ajustados de modo a minimizar o erro da rede.

A minimização do erro no algoritmo *backpropagation* é obtida pela execução do *gradiente descendente* (Gama et al. 2012), na superfície de erros do espaço de pesos, onde a altura para qualquer ponto no espaço de pesos corresponde à medida do erro. Para utilizar este algoritmo a função de ativação necessita de ser contínua, diferenciável e, de preferência, não decrescente, como a função sigmoide (Gama et al. 2012). O ajustamento dos pesos inicia-se nas unidades de saída, onde a medida do erro é conhecida e, procede com a retro propagação desse erro entre as camadas, estimando o erro dos neurónios das camadas intermédias, ajustando os pesos até que a camada das unidades da primeira camada intermédia seja processada.

Na realidade, os pesos para um neurónio particular devem ser ajustados na proporção direta do erro da unidade de processamento ao qual está conectado. Por esse motivo a retro propagação dos erros através da rede permite um ajustamento correto dos pesos sinápticos entre as camadas do modelo conexionista. Deste modo, é possível identificar duas fases distintas no processo de aprendizagem do algoritmo *backpropagation*: a primeira onde as entradas se propagam entre as camadas da rede, da camada de entrada até a camada de saída, e a segunda em que os erros são propagados na direção contrária ao fluxo de entrada.

### **3.5.3 Tamanho da Rede Neuronal Artificial**

Segundo Refenes et al. (1993) o número de unidades de processamento das camadas de entrada e saída é normalmente determinado pela aplicação. No caso das camadas ocultas, esta relação não é tão clara. O número adequado de neurónios para as camadas intermédias, segundo (Gama et al. 2012) depende de vários fatores como: o número de exemplos de



treino, quantidade de ruído presente nos exemplos, complexidade da função a ser aprendida e, distribuição estatística dos dados de treino. Para (Rumelhart et al. 1986) o ideal é utilizar o menor número possível de unidades ocultas para que a generalização não fique prejudicada. Se o número de neurónios ocultos for muito grande, a rede acaba por memorizar os padrões apresentados durante o treino. Contudo, se a arquitetura das camadas ocultas possuir unidades de processamento em número inferior ao necessário, o algoritmo *backpropagation* pode não conseguir ajustar os pesos sinápticos adequadamente, impedindo a convergência para uma solução. A experiência ainda é a melhor indicação para a definição da topologia de um modelo conexionista (Surkan e Singleton, 1990). É necessário adquirir experiência em cada caso específico pois as condições são sistematicamente diferentes, a escolha da melhor arquitetura não se apresenta como uma tarefa simples. Geralmente esta escolha é realizada recorrendo a um processo de tentativa e erro, avaliando diferentes configurações antes de decidir qual delas utilizar.

### **3.6 Árvores de decisão**

Árvores de decisão são modelos estatísticos que utilizam o treino supervisionado para a classificação e predição do resultado (Russel e Norvig, 2003). Na sua construção é utilizado um conjunto de treino, formado por entradas e saídas, ou seja elementos previamente classificados. Sendo estas últimas as classes. Um problema complexo é decomposto em sub-problemas mais simples, esta técnica é aplicada recursivamente a cada sub-problema. A capacidade de discriminação de uma árvore decorre da divisão do espaço definido pelos atributos em subespaços, associando uma classe a cada subespaço (Rich e Knight, 1991). As árvores de decisão foram aplicadas com sucesso em generalização cartográfica, ver por exemplo Plazanet et al. (1998) ou Mustière (2005).

Na construção da árvore de decisão, os dados são divididos em dois subconjuntos, de modo que os registos dentro de cada subconjunto sejam mais homogéneos do que no subconjunto precedente. É um processo recursivo, em que cada um dos dois subconjuntos é subdividido outra vez, repetindo-se este processo até que o critério da homogeneidade seja alcançado, ou até que um outro critério de paragem seja satisfeito. O mesmo campo do preditor pode ser usado diversas vezes em diferentes níveis na árvore (Rocha et al. 2008). A árvore de decisão classifica amostras, também chamadas exemplos ou padrões num número finito de classes. O sistema mais popular é o ID3 (*Iterative Dichotomiser 3*) (Quinlan, 1986). O ID3 constrói

árvores de decisão a partir de um conjunto de amostras (Plazanet et al. 1998). Afim de obter e refinar os valores dos parâmetros a utilizar no algoritmo de generalização, usou-se o algoritmo C5, ver por exemplo (Larose, 2005; Feng e A-Xing, 2003), para a aquisição do conhecimento, um melhoramento do algoritmo ID3.

Há requisitos que se deve ter em conta antes de aplicar algoritmos de árvores de decisão:

1. Os algoritmos de árvores de decisão representam aprendizagem supervisionada, como tal requer variáveis alvo pré classificadas. Um conjunto de dados de treino deve ser trabalhado, o qual fornece ao algoritmo os valores da variável alvo.
2. Este conjunto de dados de treino deve ser rico e variado, fornecendo ao algoritmo dados dos tipos de registos para cada classificação, que serão necessárias no futuro. Árvores de decisão aprendem pelo exemplo. Se porventura os exemplos têm sistematicamente faltas de subconjuntos de registos, a classificação e predição para este subconjunto poderá ser problemático ou impossível.
3. As classes de atributos alvo deve ser discreta. Ou seja, não se podem aplicar árvores de decisão a uma variável alvo contínua. Também, a variável alvo deve tomar valores que são claramente identificados como pertencendo, ou não, a uma classe particular.

As árvores de decisão procuram criar um conjunto de nós folha que são tão puros quanto possível, ou seja, onde cada registo num nó folha particular tem a mesma classificação. Desta forma, a árvore de decisão pode fornecer atribuições de classificação com a medição mais elevada da confiança disponível.

### **3.6.1 Representação de uma árvore de decisão**

Uma árvore é um conjunto finito de elementos, chamados nós ou nódulos, onde se distingue, a raiz, na qual têm origem todos os outros nós, chamados filhos. Ao nó terminal da árvore, ou seja, um nó sem filhos, é chamado folha. Uma árvore de decisão usa a estratégia de dividir para conquistar, para resolver um problema de decisão. Um problema complexo é dividido em problemas mais simples, aplicando recursivamente a mesma estratégia. As árvores de decisão classificam exemplos ou padrões ordenando-os na árvore, da raiz até algum nó folha, ver figura 3.6, que nos dá a classificação desse exemplo. Um exemplo é classificado começando no nó raiz da árvore, testando o atributo especificado por este nó, baixando a seguir no ramo da árvore que corresponde ao valor do atributo no exemplo dado. Este processo é então repetido para a sub árvore tendo como raiz o novo nó. Cada nó interno da

árvore corresponde a um teste do valor de uma das propriedades e, os ramos deste nó são identificados com os possíveis valores do teste.

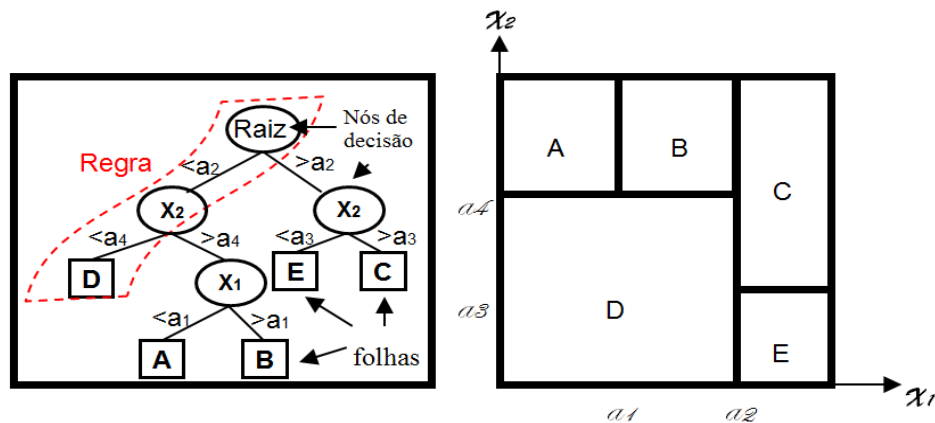


Figura 3.6 – Representação de uma árvore de decisão e respectiva representação no espaço (adaptado de Gama et al. 2012)

A figura 3.6 representa uma árvore de decisão onde cada nó de decisão contém um teste para um atributo, cada ramo descendente corresponde a um possível valor deste atributo, o conjunto de ramos são distintos, cada folha está associada a uma classe. Cada percurso da árvore, da raiz até à folha, corresponde uma regra de classificação. No espaço definido pelos atributos, cada folha corresponde a um hiper-rectângulo, ver figura 3.6.

O critério utilizado para realizar as partições é o da utilidade do atributo para a classificação. Algumas possibilidades para escolher esse atributo são:

- aleatória: seleciona qualquer atributo aleatoriamente;
- menos valores: seleciona o atributo com a menor quantidade de valores possíveis;
- mais valores: seleciona o atributo com a maior quantidade de valores possíveis;
- ganho máximo: seleciona o atributo que possui o maior ganho de informação esperado, isto é, seleciona o atributo que resultará no menor tamanho esperado das subárvores, assumindo que a raiz é o nó atual;
- razão de ganho;
- índice Gini.

Para o nosso caso aplica-se, por este critério, um determinado ganho de informação a cada atributo. O atributo escolhido como atributo teste para o corrente nó é aquele que possui o maior ganho de informação. A partir desta aplicação, inicia-se um novo processo de partição. Nos casos em que a árvore é usada para classificação, os critérios de partição mais conhecidos são baseados na entropia.

### 3.6.2 Problemas apropriados para a aprendizagem em árvores de decisão

Embora tenham sido desenvolvidos uma variedade de métodos de aprendizagem para árvores de decisão com capacidades e requisitos diferenciados, a aprendizagem em árvores de decisão geralmente ajusta-se melhor a problemas com as seguintes características (Mitchell, 1997):

- a) *As instâncias são representadas por pares de atributos-valor.* Os exemplos são descritos por um conjunto fixo de atributos (e.g. *distância*) e respectivos valores (e.g. *longe*). A situação mais fácil para a aprendizagem da árvore de decisão é quando cada atributo toma um pequeno número de valores distintos possíveis (e.g. *longe, meia distância, perto*). Contudo, as extensões ao algoritmo básico permitem utilizar também atributos com valores reais (e.g. representando numericamente a *distância*).
- b) *A função alvo tem valores discretos de saída.* Uma árvore de decisão pode atribuir uma classificação booleana (e.g. *Sim* ou *Não*) a cada exemplo. Os métodos da árvore de decisão estendem facilmente a aprendizagem de funções com mais de dois valores possíveis da saída. Uma extensão adicional permite aprendizagem de funções alvo com saídas de valores reais, embora a aplicação de árvores de decisão neste ajustamento seja menos comum.
- c) *Podem ser exigidas as descrições disjuntas.* As árvores de decisão representam naturalmente expressões disjuntas.
- d) *Os dados de treino podem conter erros.* Os métodos de aprendizagem de árvores de decisão são robustos aos erros. Tanto erros nas classificações dos exemplos de treino como erros nos valores dos atributos que descrevem estes exemplos.
- e) *Os dados de treino podem conter faltas de valores de atributos.* Os métodos da árvore de decisão podem ser usados mesmo quando alguns exemplos de treino têm valores desconhecidos, para alguns dos seus atributos.

Os problemas em que a tarefa é classificar exemplos, num conjunto discreto de possíveis categorias, são referidos frequentemente como *problemas de classificação*. A classificação será tanto melhor, quanto melhor for a qualidade dos dados dos exemplos, melhorando o desempenho futuro das árvores de decisão. Uma árvore de decisão abrange todo o espaço de instâncias ou exemplos. As regiões definidas pelas folhas são mutuamente exclusivas, ou seja, a intersecção de regiões abrangida por quaisquer duas folhas é vazia.

### 3.6.3 Regras de decisão

Um dos aspetos mais atrativos das árvores de decisão encontra-se na sua interpretabilidade, especialmente no que diz respeito à construção de regras de decisão. As regras de decisão podem ser construídas a partir de uma árvore de decisão atravessando simplesmente todo o trajeto dado do nó raiz até à folha correspondente. O conjunto completo das regras de decisão, geradas por uma árvore de decisão, é equivalente, para finalidades da classificação, à própria árvore de decisão.

As regras de decisão são constituídas da forma “SE antecedentes, ENTÃO consequente”. Para as regras de decisão, o antecedente consiste nos valores dos atributos dos ramos tomados pelo trajeto particular através da árvore, enquanto o consequente consiste no valor da classificação para a variável alvo dado pelo nó folha particular. Estas regras envolvem um atributo particular, um operador, que pode ser e.g.  $=$ ,  $\geq$ ,  $<$ ,... e um valor do domínio desse atributo. Este tipo de testes corresponde, no espaço de entrada, a um hiperplano ortogonal ao eixo do atributo testado. As regiões produzidas por estes classificadores são hiper-rectângulos, conforme figura 3.6.

O nível de confiança de determinada regra, refere a proporção de padrões ou registos no nó folha para o qual a regra de decisão é verdadeira. Se o nó folha é puro, tem como resultados níveis de confiança de 100%.

## 3.7 Árvores de classificação e regressão

CART é acrónimo de (*Classification And Regression Trees*) (Kohavi e Quinlan, 1999). Estas árvores podem ser usadas não só para classificar entidades num número de grupos discretos, mas também como uma alternativa para a análise de regressão, na qual o valor da resposta, ou seja da variável dependente, pode ser estimada. No nosso caso, a tensão a ser usada no algoritmo, é a variável dependente que pretendemos obter, como função de um conjunto de variáveis independentes, por exemplo dimensão fractal, angularidade, número de vértices. O algoritmo CART pode usar ambas as variáveis, numéricas e categóricas. Entre outras vantagens do método CART, está a sua robustez aos *outliers*.

A metodologia deste método consiste em três partes:

- 1 – Construção da árvore máxima;
- 2 – Escolha do tamanho correto da árvore;
- 3 – Classificação de novos dados usando a árvore construída.

O processo de construção começa dividindo o conjunto amostra ou (nó raiz) em nós binários, baseado numa questão simples no domínio das variáveis independentes, com uma resposta do tipo “sim” ou “não”. Inicialmente, todas as observações estão colocadas no nó raiz. Este nó é impuro ou heterogéneo, desde que contenha valores de tensões possíveis de serem usadas no algoritmo. O objetivo é arranjar uma regra, que inicialmente parte estes valores e cria grupos ou nós binários que são internamente mais homogéneos que o nó raiz.

Começando com a primeira variável, o método CART parte a variável em todos os pontos de partição possíveis. Em cada ponto de partição possível da variável, a amostra é partida em dois nós binários ou filhos. No caso de uma resposta “sim” à questão a observação é enviada para um nó e se a resposta for “não” é enviado para o outro nó. O CART aplica então o seu critério de partição para cada ponto da partição e calcula a redução na impureza, ou heterogeneidade no corte. Isto é baseado no melhor critério de corte. O objetivo é maximizar a redução no grau de heterogeneidade. Isto seleciona o melhor corte na variável, desde que para este corte a redução na impureza seja máxima. Estes passos são repetidos para cada uma das variáveis restantes, no nó raiz. O algoritmo CART, ordena então todos os “melhores” cortes para cada variável de acordo com a redução de impureza em cada corte.

Este processo seleciona a variável e o ponto onde esta variável é cortada, que mais reduz a impureza do nó raiz ou nó pai. O algoritmo CART atribui então classes a estes nós. O algoritmo CART continua o processo de corte e constrói uma árvore grande. A maior árvore que se pode realizar se o processo de corte continuar até que cada observação constitui um nó folha. Obviamente, esta árvore tem um grande número de nós terminais que são puros ou muito pequenos em conteúdo (Yohannes e Hoddinott, 1999). No entanto, nem todos os nós folha são necessariamente homogéneos, que conduz a um determinado nível de erro de classificação.

Para evitar a memorização do conjunto de treino, o algoritmo CART precisa de podar os nós e respetivos ramos descendentes, que de outra forma reduziriam a generalização dos resultados da classificação. Mesmo que a árvore completa tenha a mais baixa taxa de erro no treino ajustado, o modelo resultante pode ser demasiado complexo, tendo por resultado um sobre ajustamento aos dados. O passo seguinte é assim, um processo chamado poda da árvore, onde a árvore se torna menor, obtendo um melhor desempenho em novas situações e de uma forma mais genérica.

Enquanto o número de nós de decisão cresce, o subconjunto dos registos disponíveis para análise torna-se menor e menos representativo da população. Podar a árvore aumentará

consequentemente a generalização dos resultados. O algoritmo CART executa então a poda da árvore. Essencialmente, é encontrada uma taxa de erro total ajustada que penaliza a árvore de decisão para ter nós folha demais e assim demasiada complexidade.

### **3.8 Considerações sobre os algoritmos C4.5 e CART**

O algoritmo C4.5 é uma extensão de Quinlan do seu próprio algoritmo ID3 (Quinlan, 1986), antecedentes do C5, para gerar árvores de decisão. Estes algoritmos visitam recursivamente cada nó de decisão, selecionando a partição ótima, até que não seja possível partições adicionais. No entanto, há algumas diferenças entre o CART e o C4.5. Geralmente, os algoritmos exploram heurísticas que executam localmente uma procura que olha para a frente um passo. Uma vez tomada uma decisão, esta nunca é reconsiderada. Este tipo de procura, baseia-se numa filosofia conhecida como subida de encosta, do inglês *hill-climbing*, sem retrocesso, sendo suscetível aos riscos usuais de convergência para uma solução que é localmente ótima, mas não globalmente (Gama et al. 2012).

Contrariamente ao CART, o algoritmo C4.5 não se restringe às separações binárias. Considerando que o CART produz sempre uma árvore binária, o C4.5 produz uma árvore de uma forma mais variável. Para os atributos categóricos, o C4.5 produz à revelia um ramo separado para cada valor do atributo categórico. Isto pode conduzir a uma maior “espessura” do que a desejada, desde que alguns valores possam ter uma baixa frequência ou possam ser associados naturalmente a outros valores.

A medição da homogeneidade do nó pelo método C4.5 é completamente diferente do método CART. O algoritmo C4.5 usa o conceito do *ganho de informação* ou da *redução da entropia*, que mede a aleatoriedade de uma variável aleatória, para selecionar a separação ótima, enquanto o CART utiliza o índice Gini para esse efeito. Enquanto o C4.5 efetua as partições das variáveis categóricas diferentemente do CART, as partições para as variáveis numéricas são similares.

### **3.9 Sobre ajustamento dos dados**

Os algoritmos indutivos, como o ID3, efetuam o crescimento de cada ramo da árvore apenas a profundidade necessária para classificar os exemplos de treino. Quando os dados forem muito ruidosos ou quando a amostra de treino é demasiado pequena a predição pode resultar

enviesada. Em qualquer destes casos, este algoritmo pode produzir árvores com sobreajustamento dos exemplos de treino.

Diz-se que uma árvore se sobreajusta aos exemplos de treino, se alguma outra árvore se ajustar pior e tiver melhor desempenho sobre a distribuição completa dos exemplos, isto é, incluindo exemplos além do treino ajustado. Quando os exemplos de treino contêm erros aleatórios ou ruído, é possível uma árvore ajustar-se aos exemplos de treino melhor do que outra, mas ter um pior desempenho sobre os exemplos subsequentes.

Segundo Mitchell (1997), o sobreajustamento é possível mesmo quando os dados do treino não contêm ruído, especialmente quando um pequeno número de exemplos são associados aos nós folha. Neste caso, é possível que algum atributo divida muito bem os exemplos, apesar de não estar relacionado com a função real do alvo. Sempre que tais regularidades coincidentes existam, há um risco real de sobreajustamento. Há diversas abordagens para evitar o sobreajustamento na aprendizagem de árvores de decisão. Estas podem ser agrupadas em duas classes: a primeira refere-se a abordagens que param o crescimento da árvore mais cedo, antes que alcance o ponto onde classifica perfeitamente os dados de treino, a segunda engloba abordagens que permitem à árvore o sobreajustamento dos dados e, depois podar à posteriori. Embora a primeira destas abordagens pudesse parecer mais direta, a segunda abordagem de pós-poda de árvores sobreajustadas conclui-se ser mais bem sucedida na prática. Isto é devido à dificuldade na primeira abordagem de estimar precisamente quando parar o crescimento da árvore.

Não obstante se o tamanho correto da árvore for encontrado parando antecipadamente o crescimento ou recorrendo à pós-poda, uma pergunta que deve ser colocada é, qual o critério que deve ser usado para determinar o tamanho final correto da árvore? Várias estratégias têm sido propostas incluindo:

- O uso de um conjunto separado de exemplos, distintos dos exemplos de treino, para avaliar a utilidade dos nós, de pós-poda da árvore.
- Usar todos os dados disponíveis para o treino, mas aplicando um teste estatístico para estimar se se deve expandir, ou podar, um nó particular. Segundo (Mitchell, 1997), Quinlan (1986) usou um teste do qui-quadrado para estimar se expandindo um nó seria provável melhorar o desempenho da árvore sobre todos os dados disponíveis, ou apenas na amostra atual dos dados de treino.



Nesta abordagem, os dados disponíveis são separados em dois conjuntos: um conjunto de treino, que é usado na aprendizagem, e um outro conjunto separado, de validação, que é usado na avaliação da precisão/exatidão da árvore sobre dados subsequentes e, em particular, para avaliar o impacto de podar esta árvore.

Mesmo que quem construa a árvore possa ser enganado por erros aleatórios e por regularidades coincidentes no conjunto de treino, o conjunto de validação é, em princípio, pouco suscetível de conter as mesmas flutuações aleatórias. Consequentemente, pretende-se que o conjunto de validação forneça uma verificação de segurança para o sobre ajustamento do conjunto de treino. Naturalmente, é importante que a validação ajustada seja suficientemente abrangente para fornecer uma amostra estatística significativa dos exemplos. Uma prática heurística comum, quando não se dispõe de grandes conjuntos de dados, é reter um terço dos exemplos disponíveis para o conjunto de validação, usando os outros dois terços para o treino (Larose, 2005).

### **3.10 O Agente**

Um Agente é toda a entidade capaz de interagir com o ambiente, em geral guiado por objetivos, mas não necessariamente (Costa e Simões 2004), ou seja, poderá não ter um objetivo concreto. Em termos estruturais um agente é composto por mecanismos que lhe permitem recolher informação sobre o ambiente (percepção), mecanismos que lhe permitem atuar sobre o ambiente (ação) e processos que lhe permitem definir qual a melhor ação a realizar (decisão). Os processos de decisão serão tanto mais sofisticados quanto mais complexa for a tarefa e/ou o ambiente. Por exemplo os agentes puramente reativos são máquinas simples, sem estado interno e que se limitam a reagir aos estímulos que recebem do ambiente. Os agentes comunicam entre si podendo entrar num processo de negociação. Este processo de negociação engloba diversos mecanismos, desde leilões, protocolos baseados no Protocolo da rede de Contrato, argumentação, teoria dos jogos e aproximações heurísticas.

Apesar da sua simplicidade, os leilões são uma ferramenta poderosa que os agentes automáticos podem usar para alocar bens, tarefas e recursos (Wooldridge, 2002). Em abstrato, um leilão é executado entre um agente conhecido como o leiloeiro e um conjunto de agentes denominados licitantes. O objetivo de um leilão é que o leiloeiro consiga alocar o bem a um dos licitantes. O agente que lícita com o maior valor é aquele a quem o bem é alocado. Estes protocolos são conhecidos como leilões de primeiro preço. Outra possibilidade

é alocar o bem ao agente que efetua a licitação mais elevada, mas este agente paga somente o preço da segunda licitação. Estes leilões são conhecidos por leilões de segundo preço. A possibilidade mais simples é haver uma única oportunidade de licitação, depois do qual o leiloeiro aloca o bem ao vencedor. Estes leilões são conhecidos como lance único, do inglês “*one shot*”. Os leilões onde o preço inicial é baixo e as sucessivas licitações vão aumentando, são conhecidos por ascendentes. O outro tipo, descendente, no qual o leiloeiro começa com um valor elevado e vai descendo o preço em *rounds* sucessivos, até haver uma licitação. Há diversos tipos de leilões:

**- Leilão Inglês**

O leilão Inglês é do tipo primeiro preço, ascendente, quando nenhum agente está disposto a elevar a oferta ou licitação, então o bem é alocado ao agente que efetuou a licitação atual, mais elevada e paga o valor dessa licitação.

**- Leilão Holandês**

Os leilões holandeses são leilões descendentes. O leiloeiro começa por oferecer o bem com um valor artificial elevado, acima do valor previsto por qualquer licitante, o leiloeiro baixa então continuamente o preço da oferta do bem, de um pequeno valor, até que algum agente efetue uma oferta igual ao preço atual. O bem é então alocado ao agente que fez a oferta.

**- Leilão de primeiro-preço de oferta selada (*secreta*)**

O leilão de primeiro-preço de oferta selada é um exemplo de leilões de lance único, este é talvez o mais simples de todos os tipos de leilão. Os licitantes não têm conhecimento das ofertas feitas por outros agentes. Neste leilão, há um único *round*, em que os licitantes submetem ao leiloeiro uma oferta para o bem. Não há nenhum *round* subsequente, o bem é concedido ao agente que fez a oferta mais elevada e o vencedor paga o preço dessa oferta.

**- Leilão de Vickrey**

Os leilões de Vickrey são leilões *selados de segundo-preço da oferta*. Isto significa que há um único *round* de negociação, durante a qual cada licitante submete uma única oferta. Os licitantes não têm conhecimento das ofertas feitas por outros agentes. O bem é concedido ao agente que fez a oferta mais elevada; porém o preço que este agente paga, não é o preço da oferta mais elevada, mas o preço da *segunda* oferta mais elevada.

O leilão usado neste projeto é do tipo lance único, o bem é alocado ao preço mais elevado. O bem leilado é a tensão a usar na generalização da curva de nível respectiva, o vencedor será o método que para aquela curva apresenta a maior confiança, ou seja aquela que no seu processo de aprendizagem apresentou uma maior probabilidade de sucesso. Consequentemente a curva de nível será generalizada com esse valor de tensão, que é utilizada no algoritmo, ver Capítulo 5. Sobre este assunto ver (Wooldridge, 2002), (Costa e Simões, 2004), (Russel e Norvig, 2003).



# Capítulo 4

## Generalização Cartográfica de Linhas

---

### 4.1 Introdução

Neste capítulo são apresentados e discutidos os processos de generalização cartográfica implementados na cadeia produtiva do IGeoE. É também feita uma análise de algoritmos para generalização / simplificação de linhas e apresentado um algoritmo para generalizar linhas, baseado na minimização da curvatura da linha.

Primeiramente apresenta-se o processo de generalização objeto, da realidade para a BDG e caracterização das séries cartográficas envolvidas. Na segunda secção é apresentada a metodologia geral utilizada, iniciando com a generalização modelo englobada na preparação dos dados, incluindo a classificação e seleção das linhas de água, pontos de cota e curvas de nível, a representar na série 50k.

A generalização cartográfica de linhas é abordada nos pontos seguintes onde se apresentam alguns algoritmos de simplificação e suavização, com as principais medidas como critério. Após esta abordagem sumária dos algoritmos de generalização de linhas, é apresentado no ponto seguinte o algoritmo TLopes (Lopes, 2006), aqui proposto para a generalização de curvas de nível. Este algoritmo é baseado na tensão a aplicar à curva de modo a diminuir a sua curvatura. A analogia escolhida, e aquela que melhor descreve o pensamento aplicado a este algoritmo é o de um sistema de forças aplicado a um corpo elástico deformável, sendo esta deformação proporcional à tensão aplicada. No

final é feita a comparação entre as curvas de nível generalizadas manual e interativamente com este e outros algoritmos.

## **4.2 Caracterização das séries cartográficas envolvidas**

### **4.2.1 A Base de Dados Geográficos de média escala**

No esquema estático de aplicações geográficas, são especificadas as classes envolvidas no problema, juntamente com as suas representações e os seus relacionamentos. Para a execução da BDG de média escala, é efetuado, numa primeira fase, o levantamento dos objetos representados na escala 25k e respetivos atributos gráficos tais como cor, espessura, nível e tipo, sendo estes objetos posteriormente agrupados em classes. Após este levantamento são criadas as superclasses ou temas e especificadas as classes e respetivas relações, ou seja, é criado o modelo conceptual.

Atendendo à finalidade da BDG, ou seja, produção cartográfica das folhas da escala 25k e ainda a produção da escala 50k por generalização, é feito um levantamento dos objetos da série 50k por forma a estudar as tarefas de generalização efetuadas pelos cartógrafos de forma manual e semiautomática. Após este levantamento, os objetos são mapeados para as classes existentes e são ainda criadas novas classes, para abranger os objetos da escala 50k, não incluídos na série 25k

Depois dos objetos e respetivos atributos levantados, é necessário efetuar a sua simbolização. Pretende-se selecionar a série em causa e, automaticamente extrair da BDG os objetos necessários, de modo a efetuar a correspondente representação aquando da sua saída em papel. Após a criação destas classes, são selecionados os atributos para cada objeto. No caso das curvas de nível, são adicionados os atributos classificadores de linhas, utilizados nesta tese, e.g. dimensão fractal, angularidade, número de vértices. Para as linhas de água adicionaram-se os atributos necessários às operações de classificação e posterior seleção, para representação na série M782, 50k.

A BDG engloba os seguintes temas ou superclasses, para representação na série M888, escala 25k e conseqüente generalização para a M782, escala 50k, representado na figura 4.1.

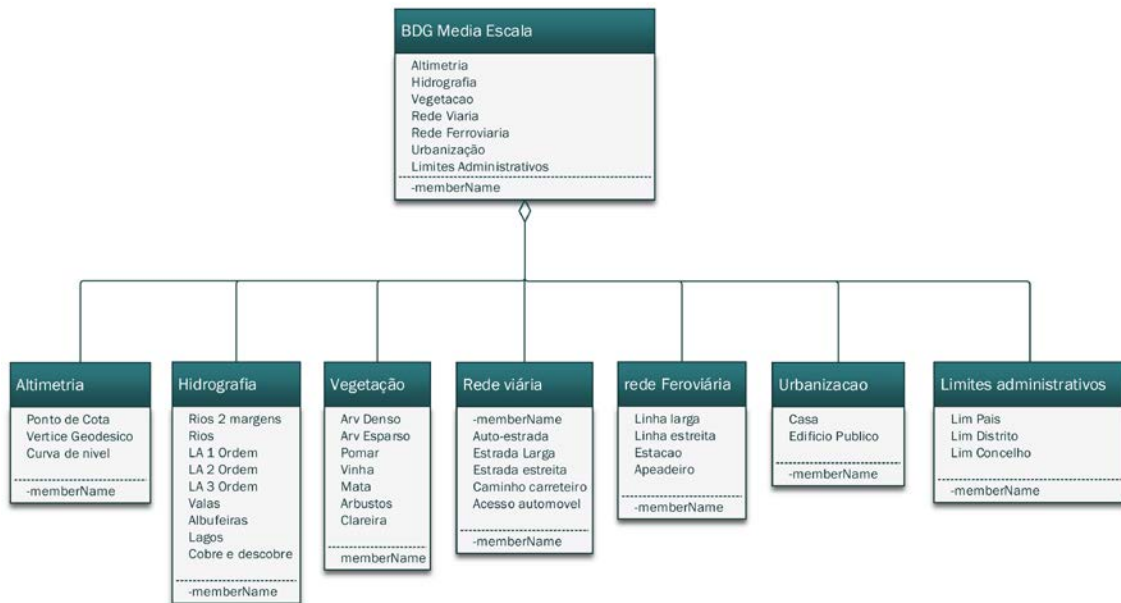


Figura 4.1 – Temas ou Superclasses representadas na série M888, generalizadas para a série M782.

#### 4.2.2 Série M888 à escala 25k

A série M888 é composta por aproximadamente 630 folhas, enquadrando o território de Portugal continental em folhas retangulares, cobrindo cada, uma área de 16km por 10km, representando as curvas de nível com uma equidistância natural de 10 metros.

A Carta militar à escala 25k é, ainda atualmente, o documento cartográfico de maior escala que cobre integralmente o território nacional, sendo por isso considerado carta base de Portugal. Os objetos representados nesta série, são obtidos por levantamento fotogramétrico, incluindo a altimetria, nomeadamente as curvas de nível, ficando o reconhecimento e completagem, a cargo da secção de topografia. Após a aquisição dos objetos em ambiente tridimensional, a carta é editada e validada. Nesta fase é colocada a toponímia, os objetos são simbolizados, contando aqui com constrangimentos de legibilidade, é também colocada a grelha de coordenadas e a legenda da carta. Posteriormente a carta é submetida a um controlo de qualidade, efetuado pela secção de controlo de qualidade que, num processo iterativo e cruzando com informação de outras fontes, verifica todos os aspetos da folha, incluindo a legibilidade. Terminado este processo a folha é dada como pronta, sendo posteriormente preparada para impressão.

O ciclo de atualização desta série varia desde os 5 anos, nas zonas de grande densidade populacional de Lisboa e Porto, 10 anos para o litoral e zonas costeiras e cerca de 20 anos para o interior do país. Desta forma, as curvas de nível não foram na sua totalidade

adquiridas com a mesma tecnologia, sendo por isso necessário ter estes factos em consideração quando se efetua a sua generalização para a série M782.

#### **4.2.3 Série M782 à escala 50k**

Esta série foi inicialmente produzida pelo IGeoE, em colaboração com o ARMY MAP SERVICE, do exército americano, sendo obtida a partir das folhas da série M888, depois de devidamente atualizadas. Segundo Mateus (1999) a generalização destas folhas era inicialmente efetuada por generalização manual, executada em estradas, linhas de alta tensão e limites administrativos. O grande problema da execução da carta da série M782, estava na atualização das folhas da série M888, escala 25k, pois quando a desatualização era grande tornava-se incomportável, inviabilizando a execução da carta M782, escala 50k (Mateus, 1999).

A equidistância natural nesta série cartográfica é de 20 metros, significando que a seleção engloba cerca de metade das curvas de nível representadas na carta 25k. O enquadramento da série 25k serve de base à série 50k, sendo cada folha desta série composta por 4 folhas 25k adjacentes, num normativo previamente fixado.

### **4.3 Generalização modelo**

#### **4.3.1 Seleção das curvas de nível**

A generalização de curvas de nível, tem início com a seleção das mesmas a partir da BDG de média escala. A equidistância natural nesta BDG é de 10 metros. A seleção efetuada tem por objetivo a representação das curvas de nível na carta da série 50k, cuja equidistância natural é de 20 metros. Desta forma apenas são selecionadas as curvas de nível cujas cotas sejam múltiplas de 20, sendo as curvas de nível mestras as curvas com cotas múltiplas de 100. As curvas de nível mestras são representadas com um traço de maior espessura, sendo interrompidas para inscrever o valor da sua cota e em casos específicos, como intersecção com alguns objetos, otimizando a legibilidade da carta. O uso de curvas de nível permite uma mais rápida avaliação do terreno pelo utilizador e uma melhor legibilidade.



### 4.3.2 Seleção dos pontos de cota

Os critérios para a seleção dos pontos de cota a representar na série 50k, têm sofrido alterações ao longo dos anos. Recentemente as normas foram alteradas, constando então que “*em zonas planas, os pontos cotados, além de definirem pontos dominantes que eventualmente existam, disseminar-se-ão numa rede aproximada de 1 km vezes 1 km*” (Normas IGeoE, 2001). Esta regra é uma orientação para o cartógrafo que poderá selecionar os pontos que considere importantes para serem representados e apagar os de menor importância, desta forma a seleção é subjetiva e totalmente dependente do cartógrafo. Contudo os pontos deviam ser generalizados, selecionando os de cota mais elevada e eliminando pontos cotados que estivessem a uma distância igual ou inferior a 500 metros de vértices geodésicos ou de outros pontos cotados mais elevados. Os vértices geodésicos devem ser selecionados de acordo com as normas vigentes.

Com estas regras, adoptou-se a estratégia de selecionar todos os Vértices Geodésicos e não selecionar os pontos cotados que estivessem no interior de um *buffer* de raio 500m com centro nestes vértices. Após esta seleção ordenaram-se, por ordem decrescente do valor da sua cota, os restantes pontos cotados e, serviram como centro do *buffer* os de maior cota apagando os que se encontravam no seu interior, percorrendo toda a lista ordenada. No final desta tarefa, o cartógrafo tem sempre a possibilidade de alterar esta seleção, poderia selecionar outros pontos que considerasse importantes ou apagar pontos selecionados, se considerasse que não deveriam ser representados.

Com a necessidade de um modelo digital do terreno detalhado, para a ortorectificação de ortofotomapas e outros produtos e aplicações, o IGeoE adoptou a estratégia de adquirir uma densidade mais elevada de pontos de cota, havendo neste caso a necessidade de uma mais apurada seleção dos pontos, para a representação na série 25k e consequentemente a sua representação na série 50k. Sempre houve a preocupação da sincronização destas duas séries, pelo que após a generalização da altimetria, o operador confirma sempre se os pontos cotados representados na série 50k, estão também representados na série 25k. Este passo é necessário pois a execução destas duas séries é geralmente simultânea, pelo menos parcialmente.

Desta forma a estratégia seguida pelos cartógrafos, é a seleção de todos os pontos de cota e VGs (*Vértices Geodésicos*) que estão na BDG e, posteriormente apagar alguns de



permanente ao longo de todo o ano. As linhas de água da folha da série 25k são classificadas em 3 ordens, sendo a 1ª ordem a mais importante e por isso representada com um traço de maior espessura. Estas ordens são classificadas de acordo com o seu comprimento. A hidrografia é representada em árvore, onde as folhas são as linhas de água de ordem mais baixa ou seja, de 3ª ordem. Além das linhas de água o tema hidrografia inclui ainda entidades areais como cursos de água, áreas de cobre e descobre e albufeiras, conforme figura 4.3.

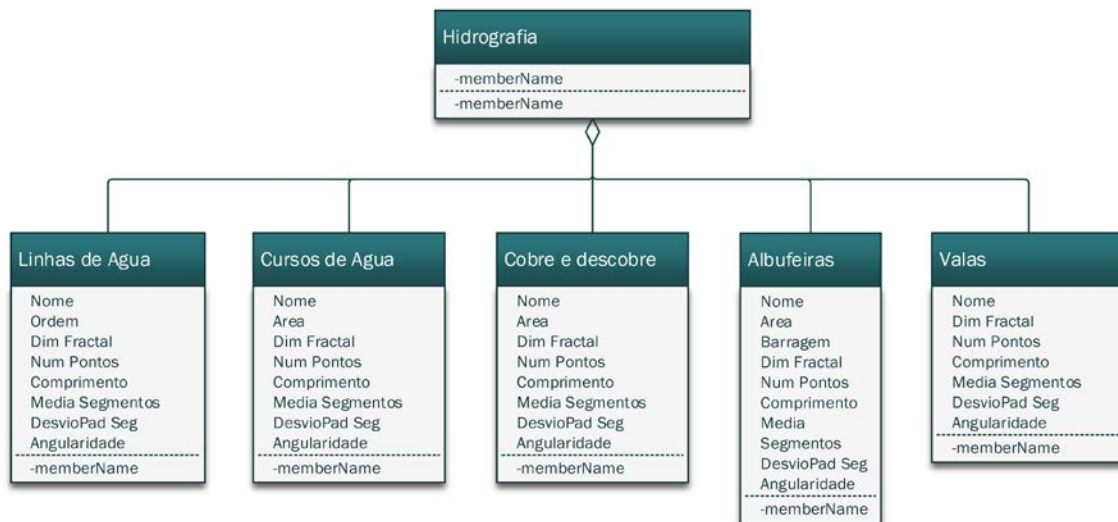


Figura 4.3 – Classes pertencentes ao tema hidrografia

Após a aquisição da rede hidrográfica na secção de fotogrametria, antes da sua inclusão na BDG, é necessário classificar estas linhas de acordo com os critérios utilizados na cadeia de produção, ou seja para representação na série 25k, trabalho este feito de forma manual, por um operador, devido à sua complexidade. No sentido de economizar recursos e uniformizar procedimentos, propõe-se a adoção da metodologia aqui apresentada, para esta classificação, carecendo contudo de verificação e finalização por parte de um operador cartógrafo.

A seleção efetuada para a série 50k, é feita a partir das linhas originais da série 25k classificadas, eliminando algumas de 3ª ordem, de modo a manter a árvore hidrográfica de uma forma legível, equilibrada e ponderada. Após esta seleção é necessário simplificar e suavizar estas linhas recorrendo a operadores e algoritmos de simplificação e suavização. As normas de produção especificam que são selecionadas para representação na série 50k as linhas de água com uma extensão superior a 1000 m. As linhas de água apenas deverão mudar de espessura nos entroncamentos com outras

linhas de água. As normas preveem ainda que não haverá distinção entre linhas de água temporárias e permanentes. As linhas de água cuja largura das margens sejam superiores a 15 metros, serão representadas como área. As linhas de água cuja largura entre margens seja não superior a 15 metros, serão representadas apenas por um traço, colapsando a área para linha.

#### **4.3.4 Classificação das linhas de água pelo método de Strahler**

O método de Strahler associa a rede hidrográfica a uma árvore com ramos e folhas em que o rio principal da rede hidrográfica é o tronco da árvore e os ramos são os afluentes. O método de Strahler classifica as linhas de água começando por atribuir o peso 1 a todas as folhas da árvore. Quando duas linhas com o mesmo peso se encontram num nó, o ramo descendente soma uma unidade ao seu peso, senão o ramo descendente fica com o peso do maior ascendente, ver figura 4.4.

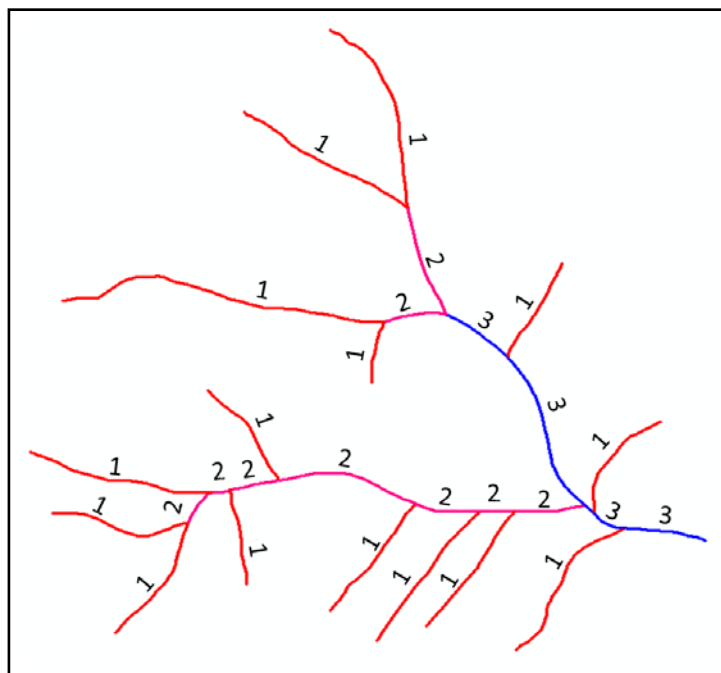


Figura 4.4 – Classificação de linhas de água pelo método de Strahler

#### **4.3.5 Classificação das linhas de água pelo método de Shreve**

O método de classificação de Shreve também começa por atribuir o peso 1 a todas as folhas da árvore, tendo o ramo descendente um peso igual à soma dos ramos ascendentes, ver figura 4.5.

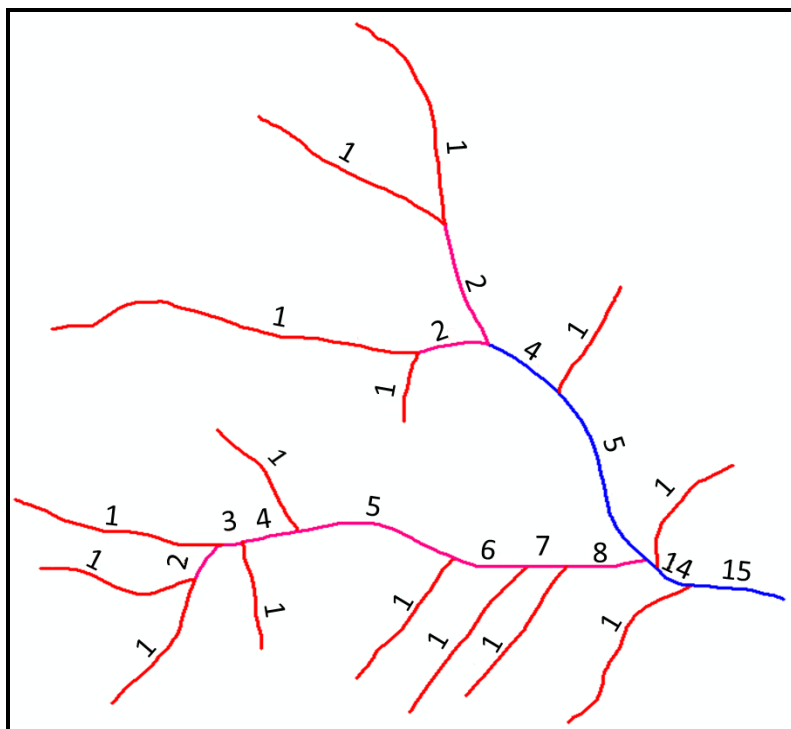


Figura 4.5 – Classificação de linhas de água pelo método de Shreve

#### 4.3.6 Classificação e seleção pelo método do IGeoE

A classificação das linhas de água representadas na série 25k e, consequente seleção para a sua representação na série 50k segue a regra geral utilizada no IGeoE e aplicada pelo operador da SCQ (*Secção de Controlo de Qualidade*) que as seleciona de acordo com o estabelecido nas normas de produção cartográfica. Na preparação dos dados é necessário que a árvore esteja segmentada nos nós. A metodologia proposta pressupõe a identificação prévia da raiz da árvore que consiste no segmento ligação a um curso de água, entidade do tipo área. Caso a folha 25k não tenha nenhum curso de água representado, então a raiz da árvore será identificada por um operador. A árvore é posteriormente construída de forma ascendente, da raiz para as folhas somando o peso 1 a cada ramo ascendente e identificando cada ramo descendente pelo seu ID (*Número Identificador*), conforme representado na figura 4.6. É importante confirmar que todos os segmentos estão ligados, para que a atribuição dos pesos seja a correta, pois a classificação posterior depende desta.

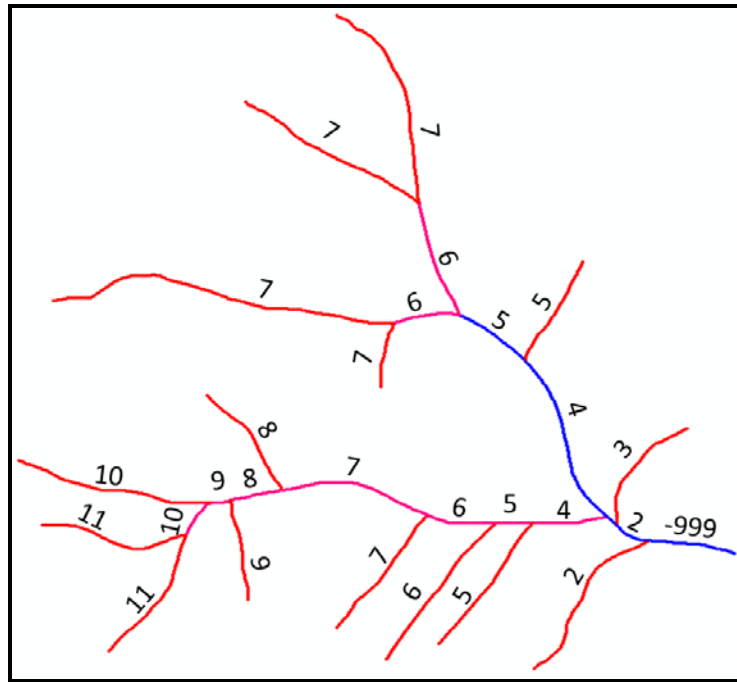


Figura 4.6 – Atribuição de pesos de forma ascendente

No passo seguinte é atribuído a ordem 3 às folhas com maior peso ascendente, sendo atribuída a ordem 3 a todos os segmentos de peso decrescente, ligados aos já classificados, até prefazer um comprimento total predefinido. Quando o comprimento total dos segmentos de ordem 3 ultrapassa o valor predefinido (1 km para a escala 25k) tem início a atribuição da ordem 2 aos segmentos descendentes. A atribuição da ordem 2 é feita até que seja atingido um comprimento total predefinido para a ordem 2 (na escala 25k o valor é de 5 km). Todos os segmentos restantes serão de classe 1.

Para os restantes ramos da árvore este processo repete-se até que os ramos encontrem outros ramos descendentes, já classificados, terminando aí o processo e passando ao seguinte com maior peso ascendente e ainda não classificado. Este processo finaliza quando todos os segmentos da árvore estiverem classificados.

Após esta classificação seleccionam-se as folhas, segmentos de 3ª ordem, com o maior peso ascendente. Será seleccionada aquela que tiver maior comprimento de entre as de 3ª ordem. Repete-se este processo para todos os ramos da árvore, até ter toda a árvore classificada e seleccionadas as folhas a representar. Na figura 4.7, as linhas de água de 3ª ordem seleccionadas têm o atributo peso, igual a 31 e são designadas por 3ª ordem 50k, pois estas serão representadas na série 50k, conjuntamente com as de 1ª e 2ª ordem.

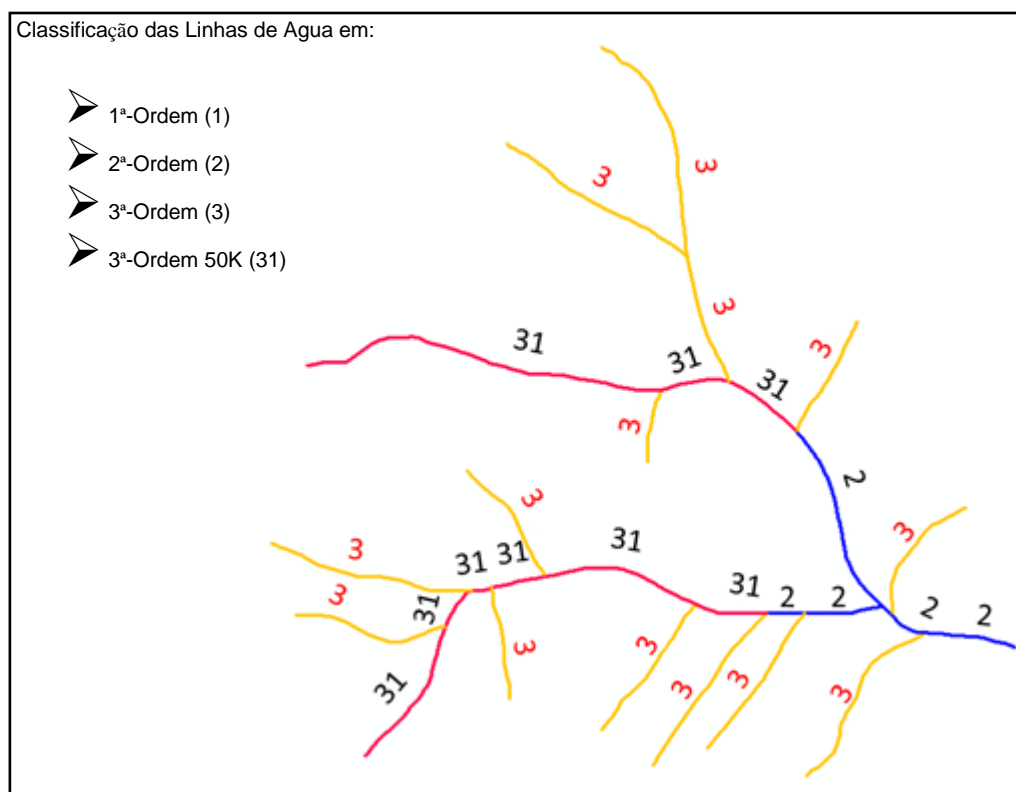


Figura 4.7 – Classificação e seleção das Linhas de água a representar na folha, escala 1/50 000

O resultado desta classificação para a folha Nº 309, escala 25k, é o apresentado na figura 4.8, reduzindo desta forma o trabalho de operador, que efetua a verificação final. Não obstante, o método apresenta alguns problemas, por exemplo junto ao limite da folha, a seleção é feita não tendo em conta os segmentos das folhas adjacentes, selecionando segmentos não adequados. O algoritmo de classificação de linhas de água demora cerca de 18 minutos para processar uma folha 25k. Claro que após esta classificação o cartógrafo terá sempre que validar ou alterar parcialmente esta classificação, podendo desta forma selecionar ramos que não seriam selecionados pelo algoritmo para representar na série 50k. Após esta seleção os cartógrafos poderão, caso seja necessário podar a árvore, ou seja ‘cortar’ as pontas das folhas das árvores, linhas de água de 3ª ordem selecionadas, em casos muito pontuais em que as pontas das folhas se tocam ou estão demasiado próximas, como por exemplo em cumes de montanhas. Outro aspeto a ter em consideração é a ligação com as folhas adjacentes, esta tarefa terá que ser supervisionada por um operador. Sobre generalização em hidrografia consultar (Sen e Gökgöz, 2012).

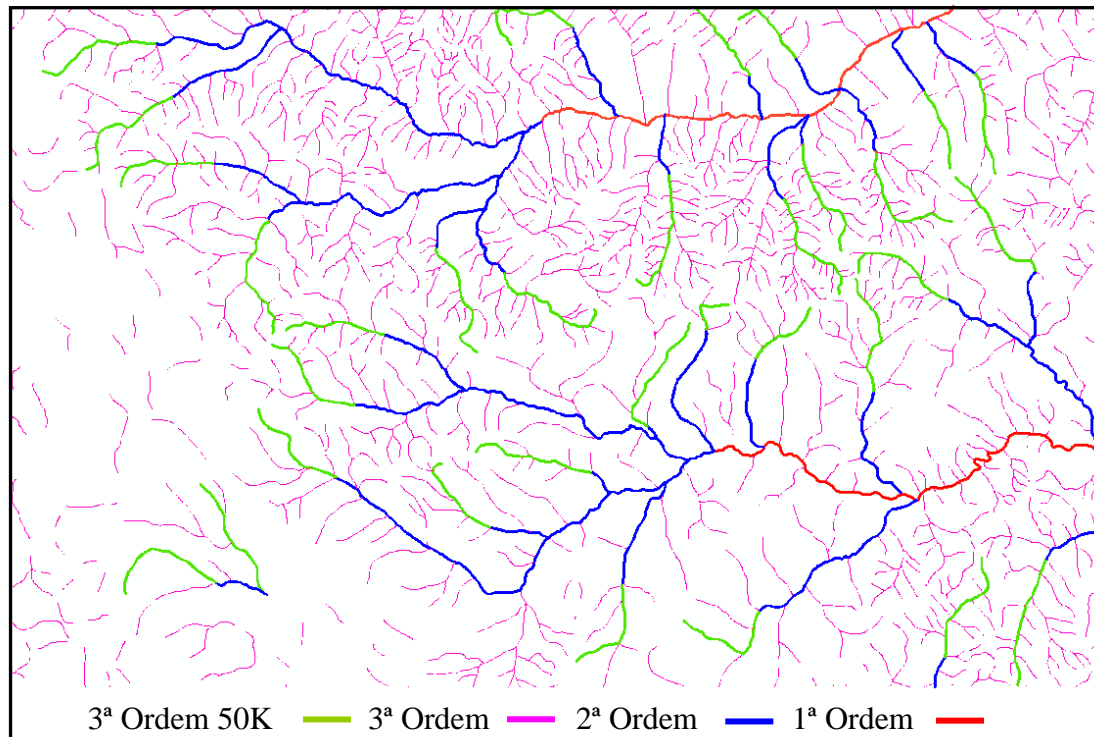


Figura 4.8 – Seleção das linhas de água a representar na folha, escala 1/50 000

#### 4.4 Algoritmos de simplificação de linhas

Os algoritmos de generalização de linhas são baseados num processo de simplificação com redução do número de pontos da linha. Esta redução de pontos na representação de uma linha, foi um assunto importante no início da era digital devido às limitações de processamento e espaço de armazenamento, como forma de reduzir a dimensão em bytes das entidades geográficas. Pretende-se com estes algoritmos, uma representação da linha original com um mínimo de pontos. Para Li (2007), não há nenhuma mudança de escala envolvida nesta operação, ou seja, a linha simplificada é para representação na mesma escala. Neste sentido é enganador utilizar estes algoritmos para propósitos de generalização ou representação multi-escala. Numa representação cartográfica da linha há pontos mais importantes para a sua representação do que outros. Os pontos mais importantes, são suficientes para caracterizar a forma do objeto. Por outras palavras, um grande número de pontos com menos informação, pode ser removido sem causar grande deformação à linha. Estes algoritmos de simplificação usam critérios geométricos de redução ou preservação de pontos, podendo ser sequenciais, iterativos ou utilizando funções dos parâmetros como critério.



#### 4.4.1 Algoritmos com parâmetros geométricos como critério

Nos algoritmos baseados no número de pontos, o  $n$ -ésimo ponto é selecionado sendo os restantes apagados. Nestes algoritmos o primeiro e o último ponto são sempre preservados. Podemos ver na figura 4.9 o caso da seleção dos terceiros e quartos pontos da linha. Outros algoritmos selecionam aleatoriamente os pontos a preservar.

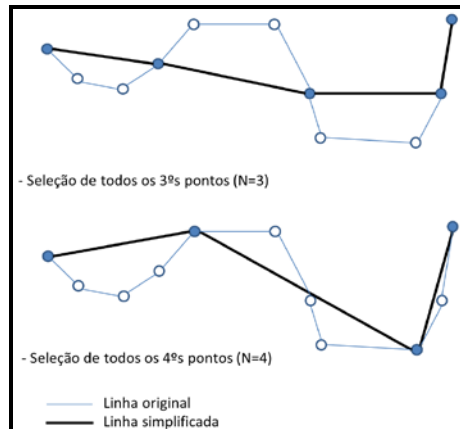


Figura 4.9 – Seleção de pontos pelo algoritmo do N-ésimo ponto

Os algoritmos baseados no comprimento usam um critério que pode ser simples ou cumulativo (soma de segmentos). Nestes algoritmos, um ponto pode ser eliminado se a distância entre eles for menor que um dado critério distância. No caso cumulativo a distância a considerar é igual à soma do comprimento de segmentos consecutivos. A distância simples refere-se à distância entre os dois pontos considerados.

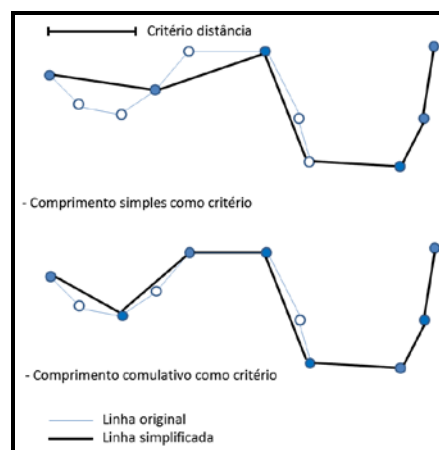


Figura 4.10 – Seleção de pontos com a distância como critério

Nos algoritmos baseados no ângulo, são preservados os pontos cujo ângulo for superior ao critério dado.

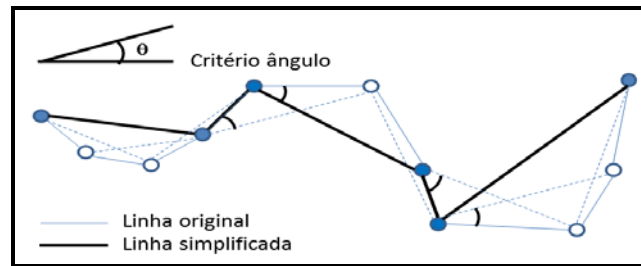


Figura 4.11 – Seleção de pontos com o ângulo como critério

O algoritmo de Reumann-Witkam (1974), começa a procura na vizinhança imediata dos pontos e avalia secções da linha usando linhas paralelas para definir a região de busca. Depois de calcular o declive inicial da região de procura, a linha é processada sequencialmente até que um dos lados do corredor de busca intersecta a linha.

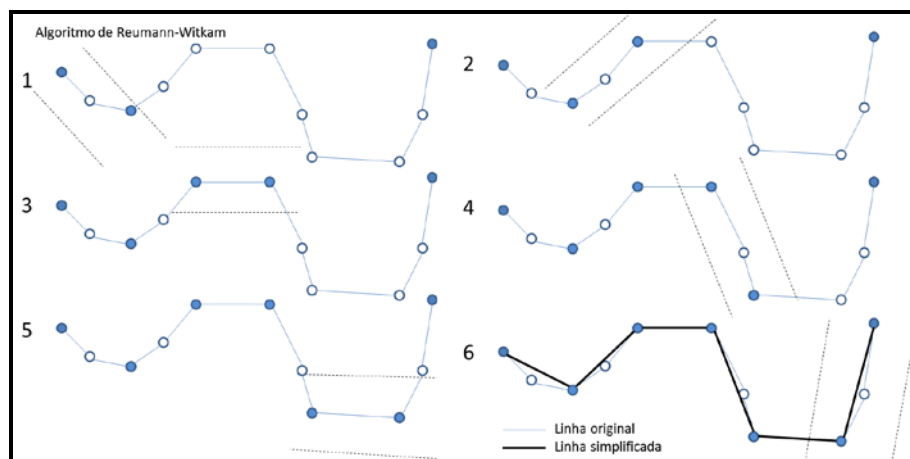


Figura 4.12 – O princípio de funcionamento do algoritmo de Reumann-Witkam

Na figura 4.12, a região de procura é recalculada baseada no último ponto de intersecção. Um ponto é intersectado onde a linha cruza a banda ou o último ponto de input contido no interior da banda é selecionado. O algoritmo continua até à utilização do último ponto e o seu arco tangente. Aqui a figura mostra todas as bandas de tolerância para a linha original. A linha final simplificada está desenhada com os pontos retidos como círculos.

Algoritmos de partição progressiva baseada na distância perpendicular. Como exemplo apresenta-se o algoritmo de Douglas-Peucker, ver figura 4.13. Este algoritmo começa por considerar uma linha base, entre o primeiro e o último vértice da linha. Posteriormente medem-se todas as distâncias perpendiculares entre os restantes vértices e a linha base. A maior distância medida, superior à tolerância, indica o vértice a ser preservado, partindo a linha em duas partes. Repete-se este procedimento para as duas

partes da linha. E assim sucessivamente até não haver vértices a uma distância superior à tolerância dada.

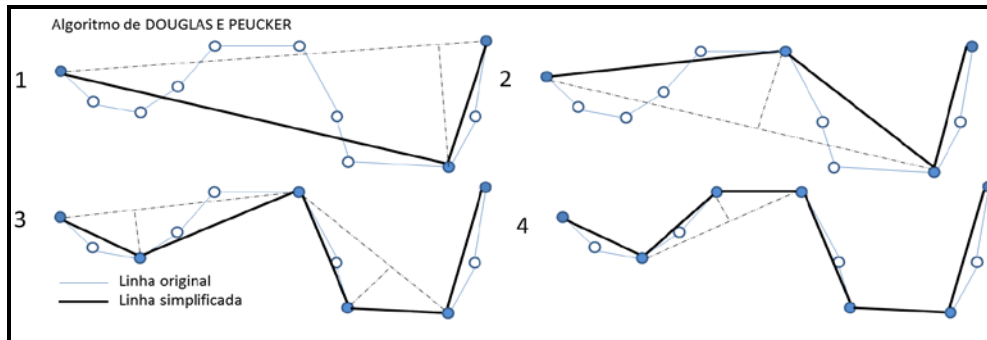


Figura 4.13 – Algoritmo de Douglas-Peucker

Algoritmos de partir e juntar (*split and merge*) baseados na distância perpendicular. Como exemplo deste tipo de algoritmos temos o algoritmo de Ansari-Delp, ver figura 4.14. Este algoritmo começa por classificar arbitrariamente um conjunto de pontos, como pontos críticos. A partir destes pontos são criadas as linhas base e medidas as distâncias perpendiculares entre os pontos da linha e estas linhas base. São depois seleccionados os pontos cujas distâncias são superiores ao critério dado.

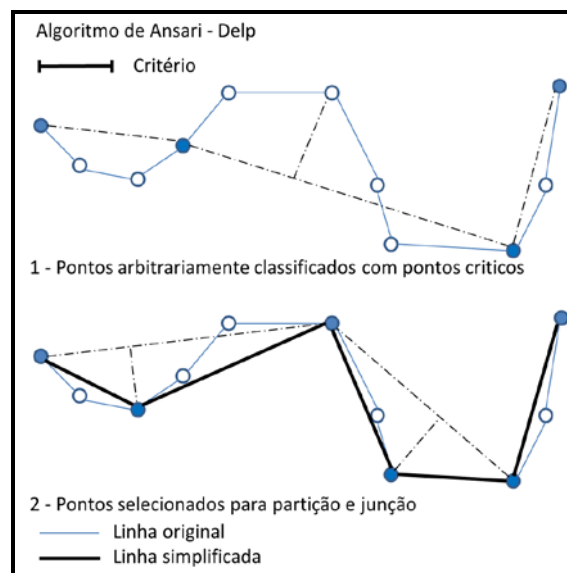


Figura 4.14 – O princípio de funcionamento do algoritmo de Ansari-Delp

Algoritmos baseados na área, e.g. algoritmo de Visvalingham-Whyatt. Se a área for superior ao critério então o ponto é preservado, ver figura 4.15. As áreas são recalculadas após apagar os pontos seleccionados.

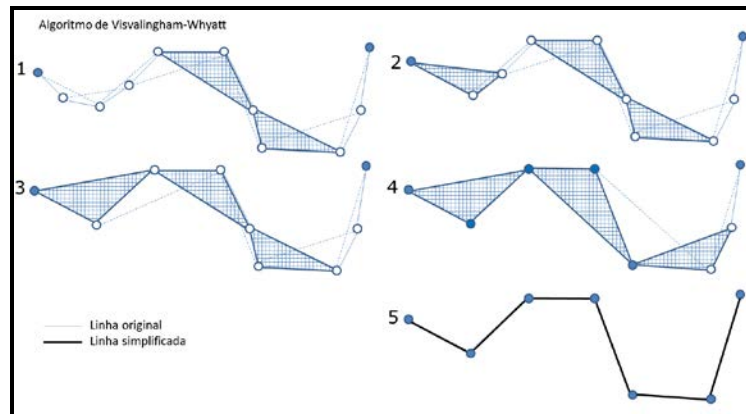


Figura 4.15 – O princípio de funcionamento do algoritmo de Visvalingham-Whyatt

#### 4.4.2 Algoritmos com funções de parâmetros geométricos como critério

Neste tipo de algoritmos utilizam-se funções matemáticas dos parâmetros geométricos dos segmentos, em vez dos próprios parâmetros. Algoritmo baseado no cosseno, como exemplo pode-se referir o algoritmo de Rosenfeld e Johnston, ver figura 4.16. Neste caso é calculado o cosseno entre os segmentos do ponto  $i$ ,  $i+k$  e  $i-k$ . Varia-se  $k$  até  $m$ , número de pontos a ter em consideração. Os pontos com o maior valor do cosseno serão preservados, como pontos críticos.

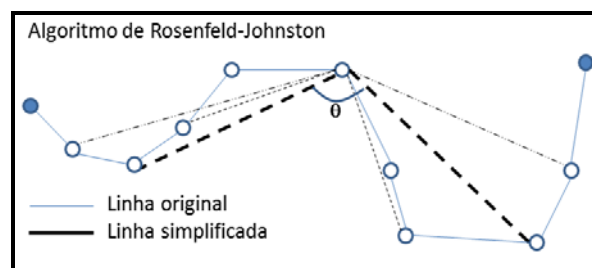


Figura 4.16 – O princípio de funcionamento do algoritmo de Rosenfeld-Johnston

Algoritmo baseado no rácio Distância/Corda, se o quociente  $d/S$  for superior a um dado valor de tolerância, então o ponto é preservado, ver figura 4.17.

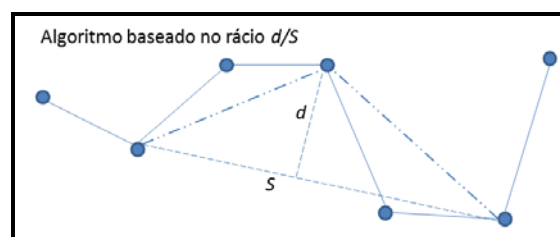


Figura 4.17 – O princípio de funcionamento do algoritmo baseado no rácio  $d/S$

Algoritmo baseado no rácio comprimento local (*LLR – local length ratio*), algoritmo de Nakos-Mitropoulos, neste caso utiliza-se uma circunferência de raio  $R$  e, são preservados os pontos cujo rácio seja superior ao valor da tolerância. Dado por:  $L/S$ , em que  $L = 2R$ , ver figura 4.18.

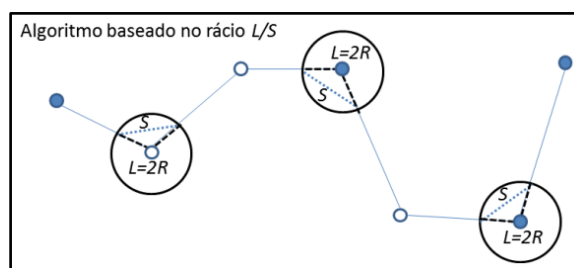


Figura 4.18 – Algoritmo de Nakos-Mitropoulos, usando o rácio  $L/S$

Existem muitos outros algoritmos de generalização de linhas, utilizando outras estratégias. Para informação adicional sobre algoritmos e respetiva classificação, ver relatório do projeto AGENT, (Shi e Cheung, 2006; Li, 2006, 2007). Apesar do elevado número de algoritmos para simplificação de linhas, há questões que é necessário ter em consideração quando se recorre a algoritmos para a generalização de linhas:

- Os pontos preservados são pontos com expressão no terreno?
- Os pontos críticos estão localizados exatamente sobre a linha?
- Há conflitos topológicos com as novas linhas, formadas pelos pontos selecionados?

Para responder a estas questões é necessário uma inspeção visual dos resultados. É necessário também uma validação geométrica quantitativa e qualitativa, com toda a carga de subjetividade inerente.

Claro que o resultado final das operações de generalização de linhas dependerá do conjunto inicial de pontos críticos. Não há garantias de resolução, no sentido de evitar as auto-intersecções ou intersecções cruzadas entre as linhas. Pequenos *spikes*, ou saliências, são sempre retidos, deste modo poderá haver uma grande distorção da forma. Os pontos selecionados não são necessariamente críticos para a representação da linha.

A redução de pontos não é um assunto fundamental na representação espacial em multi-escala (Li, 1993, 2007; Christensen, 1999, 2000). Numa tentativa de melhorar os resultados obtidos pela aplicação destes algoritmos, alguns autores propuseram

procedimentos geométricos para remover conflitos espaciais: ponto - ponto ou ponto - linha (Müller, 1990; Berg et al. 1995, 1998), outros melhoraram o algoritmo de Douglas-Peucker (Zhang e Tian, 1997).

Outras estratégias envolveram a tentativa de tornar os algoritmos auto-adaptáveis. Nestes casos os algoritmos separam a linha em troços como pré-processamento, sendo depois aplicada uma redução de pontos, com diferentes critérios para diferentes troços. De qualquer forma os pontos selecionados dependem sempre do algoritmo aplicado e dos parâmetros respetivos.

#### **4.4.3 Algoritmos de suavização de linhas**

A suavização de linhas é um processo de filtragem para reduzir as descontinuidades de 1ª ordem da linha. Esta não está diretamente relacionada com a variação da escala. A suavização pode ser efetuada no espaço ou no domínio das frequências. No domínio das frequências a ideia é basicamente, reduzir os componentes de altas frequências de uma dada linha. Os métodos mais utilizados são a transformação de Fourier e Wavelet. No domínio do espaço há duas abordagens mais usadas: médias móveis, pesadas ou não, e ajustamento de curvas. Destes ajustamentos podemos distinguir ajustamento por mínimos quadrados, ajustamento exato e ajustamento por balanceamento de energia.

As curvas usadas incluem polinómios, normalmente de 3º grau, curvas de Bézier e Splines, a figura 4.19 mostra a linha suavizada com diferentes parâmetros.

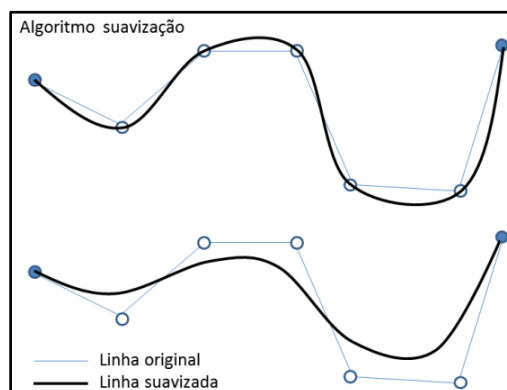


Figura 4.19 – Algoritmo de Suavização

Por norma após a aplicação de algoritmos de simplificação de linhas, segue-se a aplicação de algoritmos de suavização de linhas, com o objetivo de reduzir os vértices

angulosos, que alteram a forma das linhas nomeadamente curvas de nível. Pretende-se desta forma melhorar a legibilidade da carta e torná-la mais agradável, melhorando a sua interpretação.

#### 4.5 Algoritmo TLopes

Para a descrição do algoritmo utilizado nesta tese, algoritmo de simplificação e suavização de linhas, foi feita uma analogia com a deformação de materiais elásticos, aos quais é aplicado um sistema de forças. Considera-se a curva de nível como o limite de um corpo elástico bidimensional e isotrópico, sobre o qual atua um sistema de forças. Este corpo é deformado de modo a reduzir os máximos relativos, aproximando-os do centróide do corpo, e aumentar os mínimos relativos, afastando-os do centróide, ou seja ainda, aproximando ambos de uma linha média imaginária que seria o máximo de deformação possível, aplicável a este corpo (curva de nível). É como se houvesse por parte da curva de nível uma reação à ação efetuada pelo sistema de forças aplicado. A reação da curva de nível é notória nos pontos mínimos, visto estes serem os que mais se afastam do centróide do corpo, ver figura 4.20.

A deformação experimentada pelo corpo é contínua e proporcional à tensão aplicada, ou seja quanto maior a tensão, maior é a deformação da curva de nível.

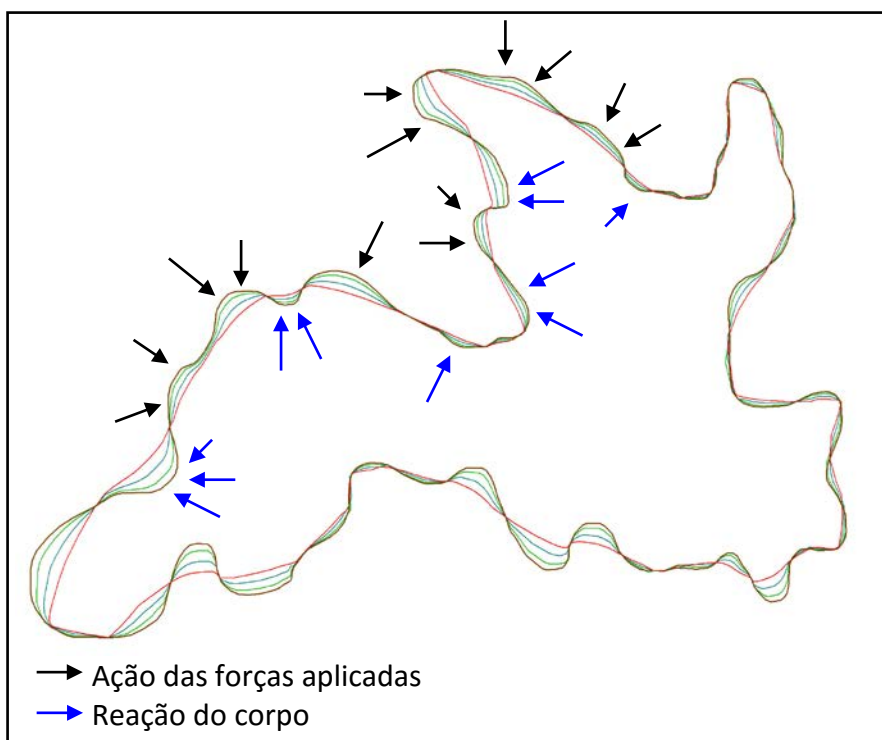
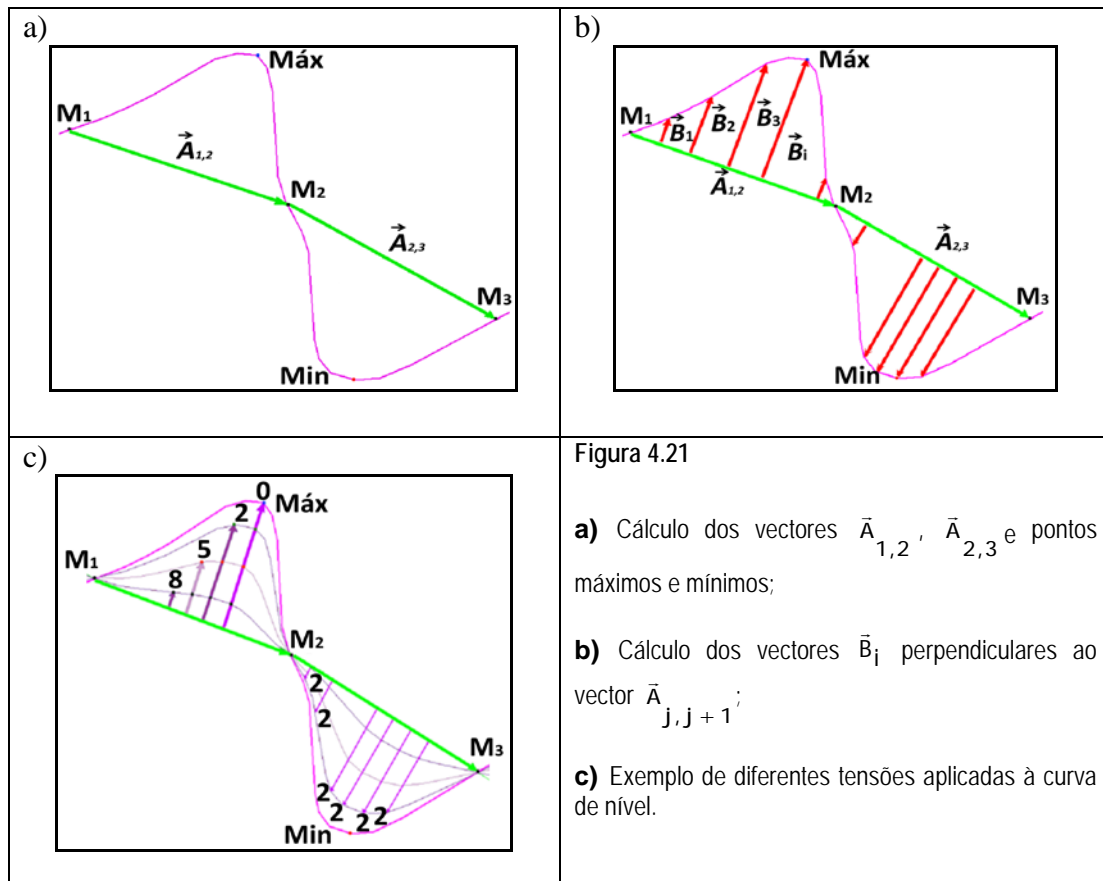


Figura 4.20 – Ação e reação da curva de nível ao sistema de forças aplicado

Seja uma curva de nível, cuja função é dada por  $y = f(x)$ , representada por uma sequência de pontos  $x_i$ ,  $i=1,...,n$ . O processo de generalização da curva de nível é realizado em 5 passos sequenciais:

1. Calculam-se os pontos máximos e mínimos relativos, onde a derivada  $\frac{df}{dx} = 0$ , ou mais precisamente onde há mudança no sinal da derivada; Incluindo os pontos (*Max* e *Min*) e ainda os pontos onde a derivada não está definida.
2. Entre cada dois pontos consecutivos calculados no ponto 1, calcular o comprimento da curva de nível entre estes dois pontos e calcular o ponto médio; ( $M_j$ ).
3. Calcular o vector  $\vec{A}_{j,j+1}$ , entre dois pontos médios consecutivos ( $\vec{A}_{j,j+1} = \overrightarrow{M_j M_{j+1}}$ ).
4. Calcular os vectores  $\vec{B}_i$ , perpendiculares ao vector  $\vec{A}_{j,j+1}$  calculado, entre este e os pontos da curva de nível.
5. De acordo com a tensão a aplicar, reduzir a norma dos vectores  $\vec{B}_i$  e, calcular as novas coordenadas dos vértices da curva de nível.





Este algoritmo foi programado em VB (Visual Basic) para o Geomedia (Intergraph), cuja *interface* se apresenta na figura 4.22. Esta aplicação pode ser utilizada interativamente pelo operador, que selecciona a curva de nível a ser generalizada, o valor da tensão e o nome da *feature class* de saída.

O valor da tensão a aplicar à curva, pode ser escolhido no slider apresentado e compreende valores entre 0 e 10. Correspondendo o zero a nenhuma deformação, pois a tensão aplicada ao corpo é nula e, a máxima deformação corresponde a 10. O operador deve ainda escolher o valor de *Kappa*, ou seja a menor distância a considerar entre pontos máximos e mínimos consecutivos. Este valor depende da razão das escalas envolvidas e é seleccionado pelo operador baseado no seu conhecimento e experiência. Por exemplo, para a generalização da escala 25k para 50k o valor recomendado é 100 m.

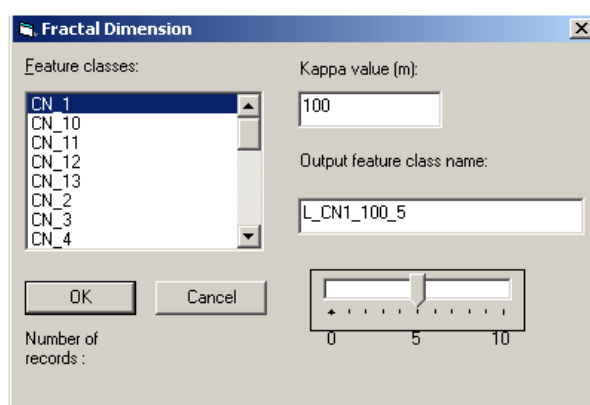


Figura 4.22 – Interface do programa para a generalização das curvas de nível.

## 4.6 Comparação com outros algoritmos de generalização de linhas

Nesta secção avaliamos a performance deste algoritmo, por comparação com dois outros algoritmos de simplificação e suavização de linhas e posteriormente com linhas generalizadas manualmente. Contudo devem ser efetuadas mais experiências para a utilização deste algoritmo noutras escalas e outros temas.

Como se pode ver nas figuras 4.23 e 4.24, com este algoritmo as linhas não necessitam de uma suavização adicional como nos algoritmos tradicionais de simplificação de linhas, em que estas apresentam pontos angulosos e bem definidos, tendo esteticamente uma forma sem aderência com a realidade. Além disso, os resultados obtidos com os

algoritmos tradicionais, demonstram uma grande sensibilidade aos valores dos parâmetros de entrada, obtendo por vezes resultados imprevisíveis, ou seja com uma pequena variação dos parâmetros os resultados podem ser muito diferentes. Só com uma grande experiência o cartógrafo pode em tempo útil obter bons resultados.

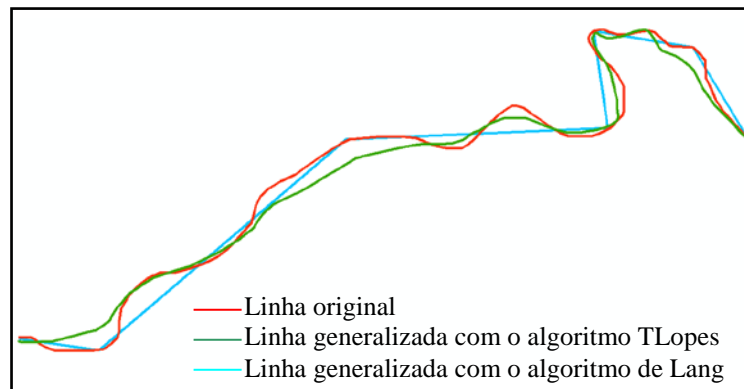


Figura 4.23 – Comparação com o algoritmo de Lang

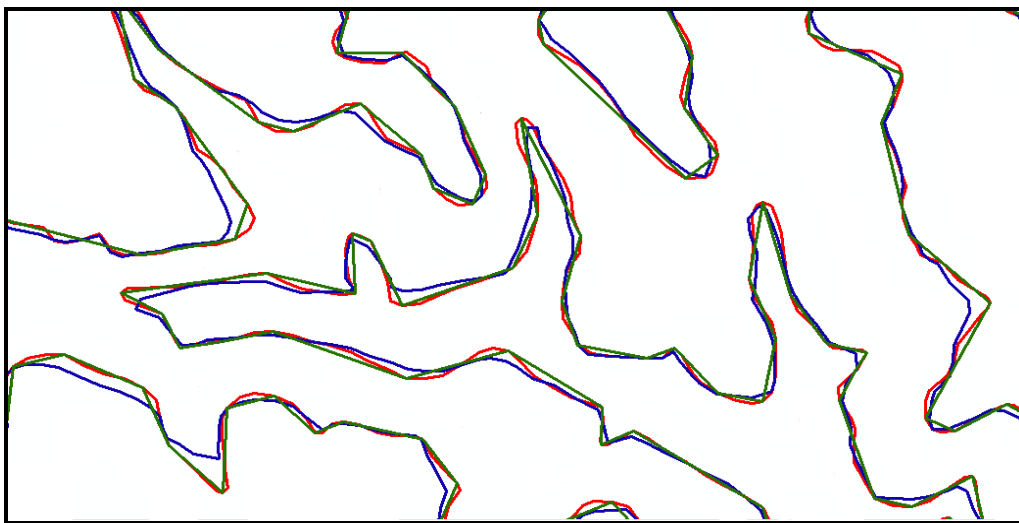
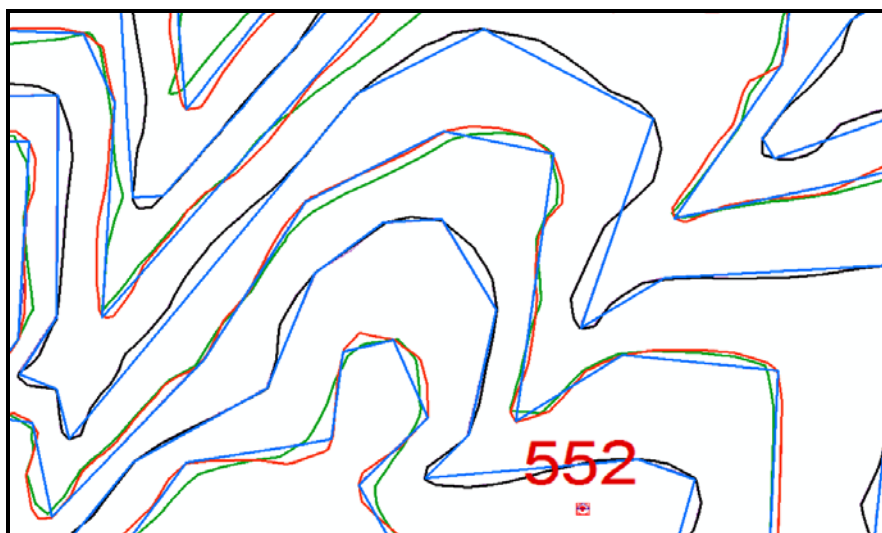


Figura 4.24 – Comparação com o algoritmo de Douglas-Peucker.

Os parâmetros utilizados nos algoritmos de simplificação de linhas aplicam-se à totalidade dos pontos da linha, não são ajustados às características locais das linhas. Por isso, alguns autores defendem uma segmentação das linhas, aplicando diferentes algoritmos e/ou parâmetros aos segmentos resultantes. Estas metodologias de segmentação de linhas não estão disponíveis em software comercial, o que dificulta a sua utilização. Além disso a comparação de resultados é feita em relação à utilização de um número finito de parâmetros, o que se adapta melhor em determinados segmentos da

linha e outros resultam em erros, nalguns casos inadmissíveis. Na figura 4.24 podemos ver que nestes segmentos os resultados podem considerar-se admissíveis para a escala 50k, tanto com a utilização do algoritmo de Douglas-Peucker com o parâmetro 10m ou com a utilização do algoritmo aqui apresentado com uma tensão de 2.



**Figura 4.25** – Comparação com o algoritmo de Douglas-Peucker. Preto – curva de nível ímpar original escala 25k, Vermelho - curva de nível par original escala 25k, Azul - curva de nível generalizada com algoritmo de Douglas-Peucker, Verde – curva de nível generalizada com algoritmo TLopes

Na figura 4.25 podemos ver que com os mesmos parâmetros as linhas generalizadas com o algoritmo de Douglas-Peucker necessitam de uma suavização adicional. Um dos problemas já evidenciados na utilização destes algoritmos é a auto-intersecção, ou seja a curva de nível após a aplicação do algoritmo pode intersectar-se a si própria. Outro problema é chamado de intersecção cruzada, após a generalização, a curva de nível generalizada intersecta outra curva de nível, criando aqui um problema de topologia, não admissível. O algoritmo proposto por Lopes (2006) propõe-se solucionar estes problemas. Além da comparação de resultados com outros algoritmos de simplificação de linhas, foi também efetuada uma comparação com as linhas generalizadas manualmente, conforme se pode ver na figura 4.26. Quando se generaliza manualmente uma curva de nível a preocupação fundamental é que o resultado seja uma linha que represente a curva que se pretende generalizar, adaptada para a escala resultante e não altere a percepção que se pretende transmitir, ou seja reduzir os detalhes e não alterar significativamente a sua forma.

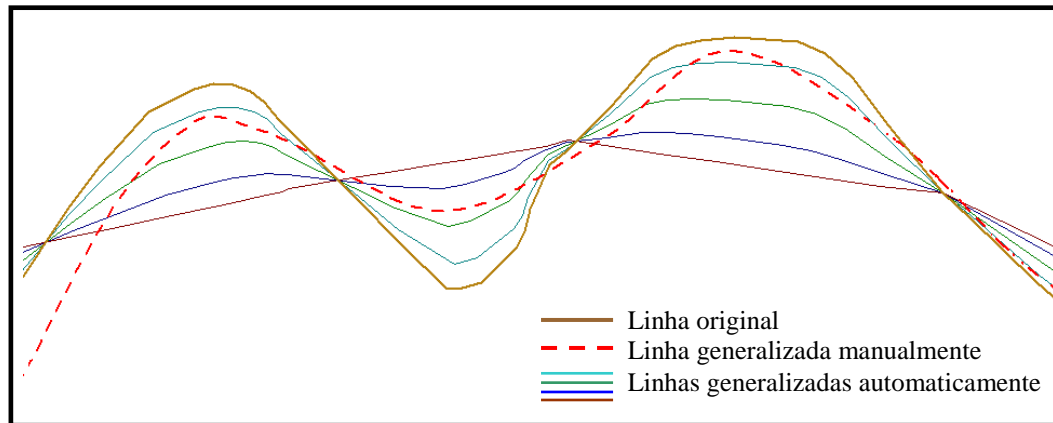


Figura 4.26 – Comparação das linhas generalizadas manualmente com as generalizadas automaticamente com o algoritmo TLoopes.

Este algoritmo aplica de uma forma expedita, os princípios da deformação à generalização de curvas de nível. Além destas comparações com o resultado obtido com outros algoritmos em generalização interativa, efetuou-se ainda o MDT (Modelo Digital do Terreno) em formato GRID (Grelha regular de pontos), conforme figura 4.27, a partir das curvas de nível originais à escala 25k, com um espaçamento de 5 metros.

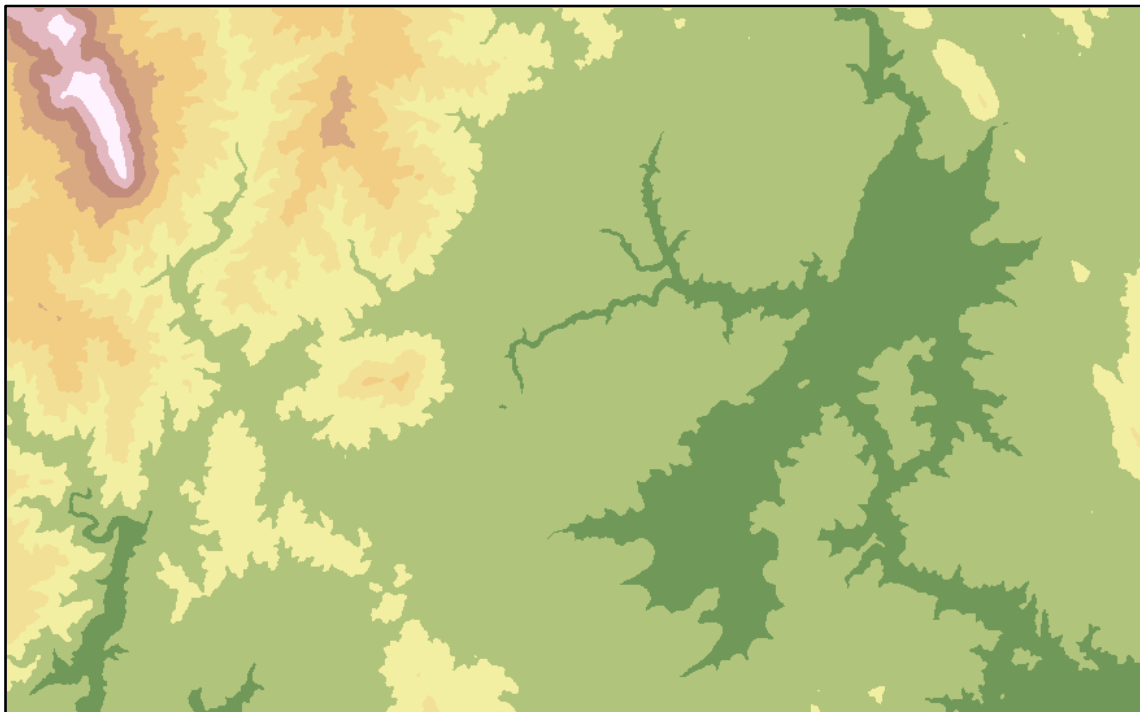
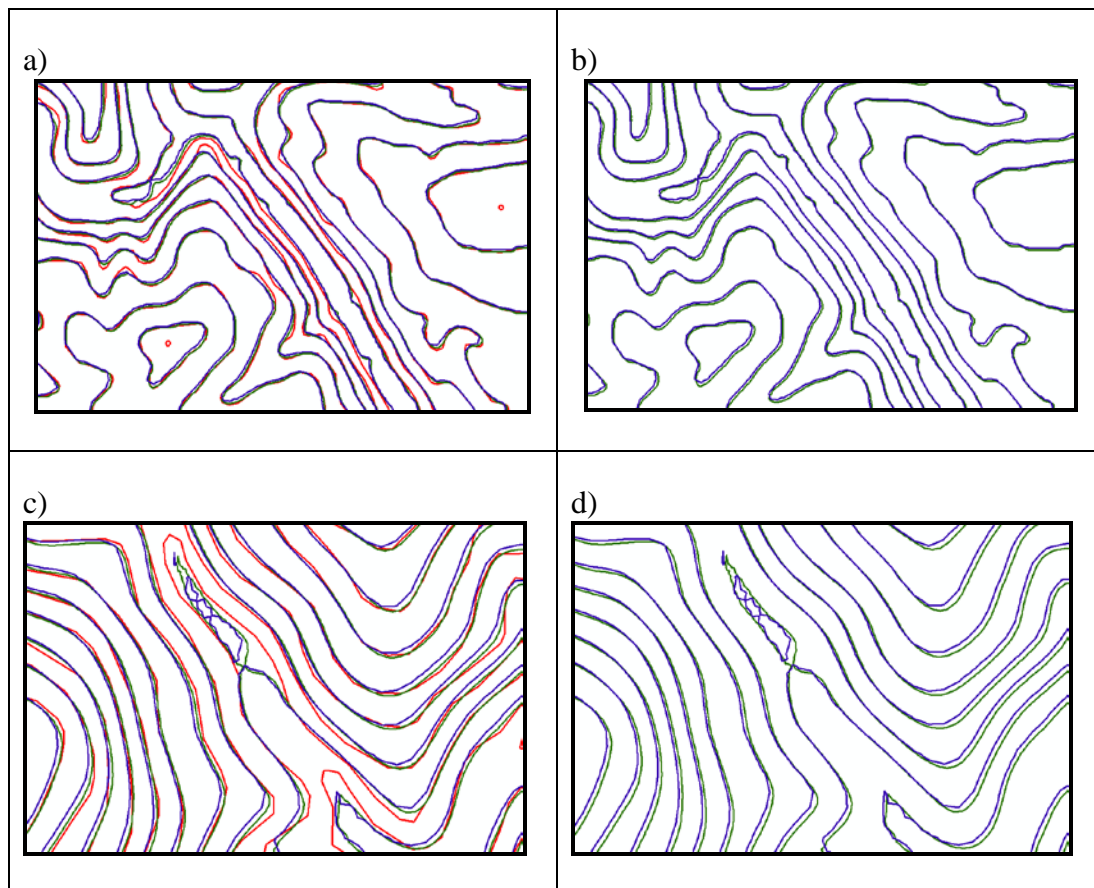


Figura 4.27 – MDT pixel de 10 metros (classes altimetria)

Como se pode constatar pela figura 4.28, a curva de nível vectorizada a partir da grid de 5m, obtida a partir das curvas de nível originais à escala 25k, não coincide exatamente

com as curvas que lhe deram origem. As curvas são coincidentes na sua generalidade, ou admissíveis, mas há casos de discrepâncias, como se pode ver na figura 4.28.



**Figura 4.28** – Extratos da folha com curvas de nível originais, escala 25k, de cor vermelha, versus curvas vectorizadas grid de 5 metros, cor verde, e vectorizadas após reamostragem para grid 10m cor azul.

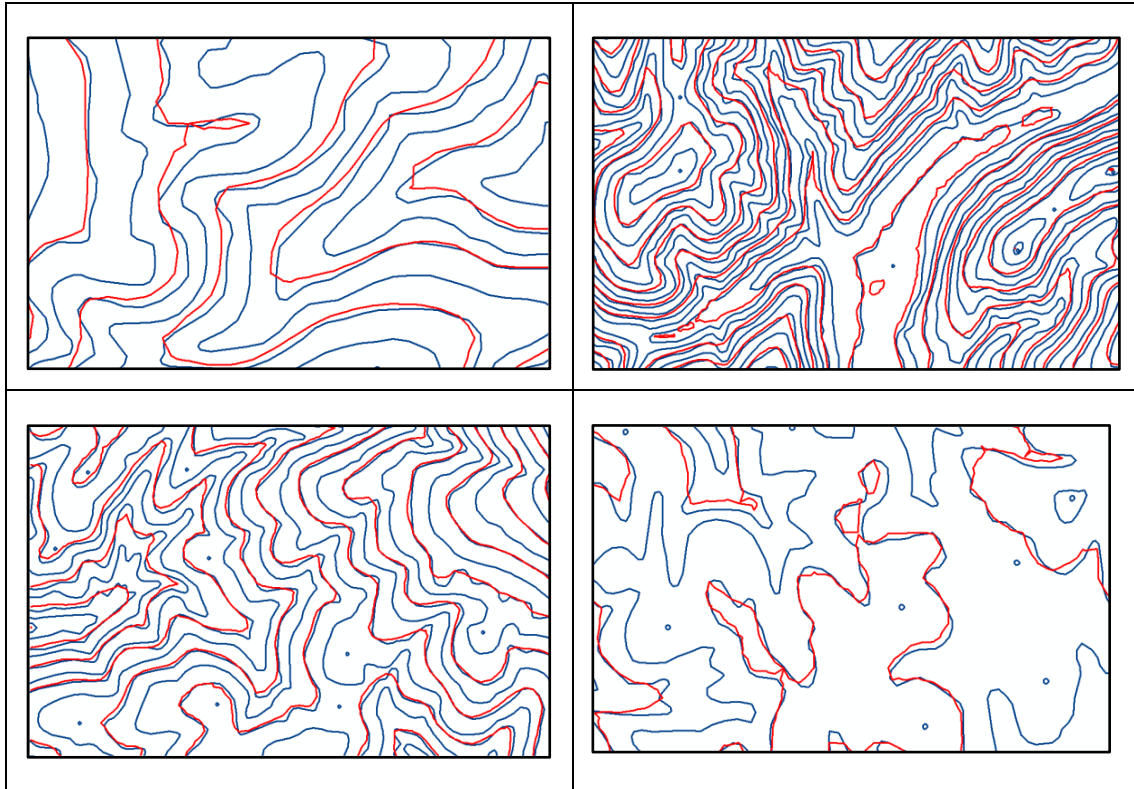
a) e c) incluindo as curvas originais;

b) e d) apenas com as curvas generalizadas.

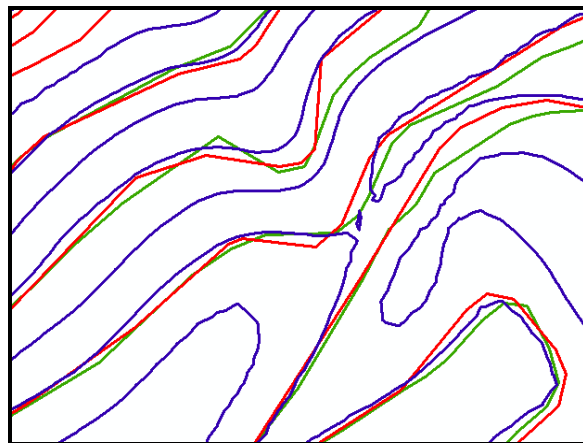
Concluimos assim que esta transformação e respetiva vectorização introduz erros, que poderão ter consequências na generalização do MDT e posterior vectorização das curvas de nível. Também a discretização do MDT em grelhas retangulares introduz erros que se irão propagar nas fases posteriores. Efetuou-se uma reamostragem deste MDT para um espaçamento de 10 metros e vectorizaram-se as curvas de nível a partir deste MDT generalizado.

Estes resultados dependem do declive do terreno, pois passamos de uma célula com tamanho de 10 metros para uma de 20 metros. Os resultados da reamostragem dependerão das características do terreno envolvido, nomeadamente o seu declive, que

afetarão diretamente as curvas de nível vectorizadas a partir deste MDT generalizado. Conforme podemos ver nas figuras seguintes as diferenças entre as curvas de nível originais e obtidas após o resampling para 20 metros.



**Figura 4.29** – Extratos da folha com curvas de nível originais, escala 25k, de cor azul, versus curvas vectorizadas, após reamostragem, de cor vermelha, generalizadas para a escala 50k.

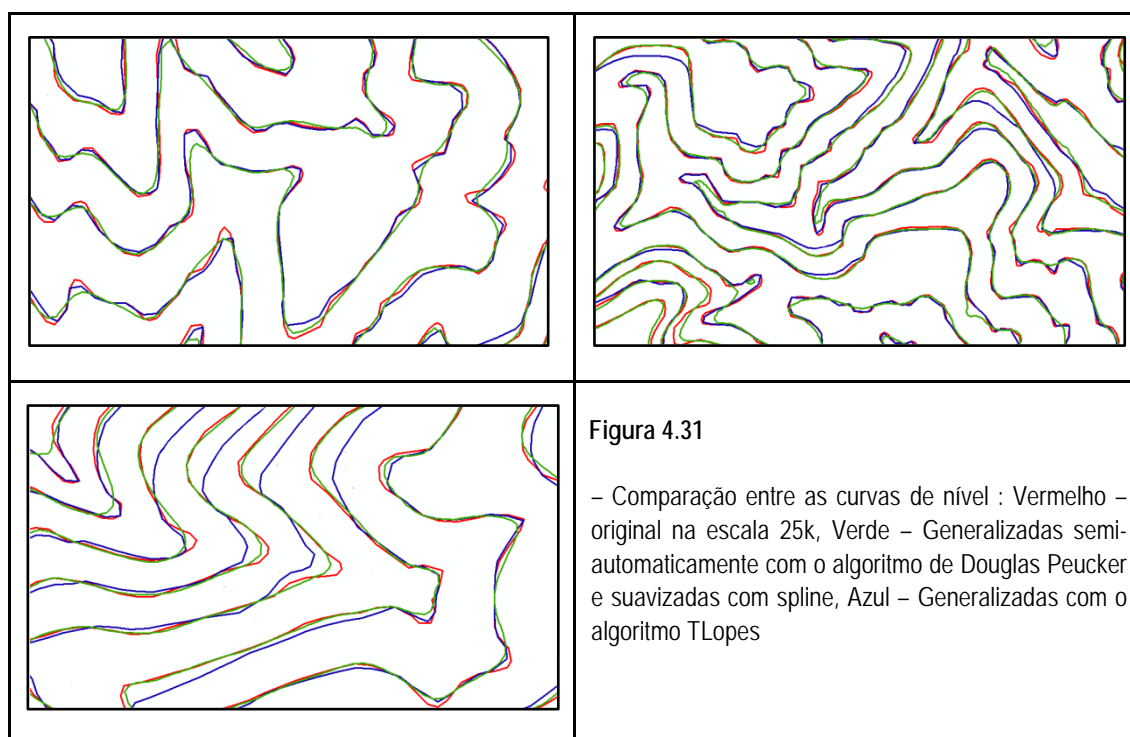


**Figura 4.30** – Comparação entre curvas generalizadas manualmente a vermelho; vectorizadas após reamostragem da grid para 10m a azul, e verde generalizadas com o algoritmo TLopes.

Como podemos ver nas figuras 4.29 e 4.30, a generalidade das linhas apresentam um resultado aceitável, havendo no entanto partes de algumas linhas com problemas



topológicos. Estas situações demonstram que ainda não é completamente dispensável o trabalho final de edição por parte do operador. Na figura 4.30 comparam-se as curvas generalizadas manualmente, representadas a vermelho, com as obtidas por vectorização do MDT generalizado, a azul e, onde são mostradas também as curvas generalizadas com o algoritmo aqui apresentado a verde. Podemos ver que há problemas topológicos nas curvas obtidas por vectorização do MDT generalizado.



Conforme podemos ver na figura 4.31, a utilização do algoritmo de Douglas-Peucker na simplificação destas curvas de nível seguido de uma suavização recorrendo a uma spline é admissível, tal como o resultado alcançado com o algoritmo aqui proposto. Este resultado é obtido em grande parte da área das folhas da 25k do IGeoE a generalizar, envolvendo algum trabalho manual para a sua finalização. Na figura 4.32 efetua-se a comparação entre as curvas generalizadas manualmente e generalizadas com o algoritmo TLopes, podemos verificar que o resultado é admissível na generalidade da folha.

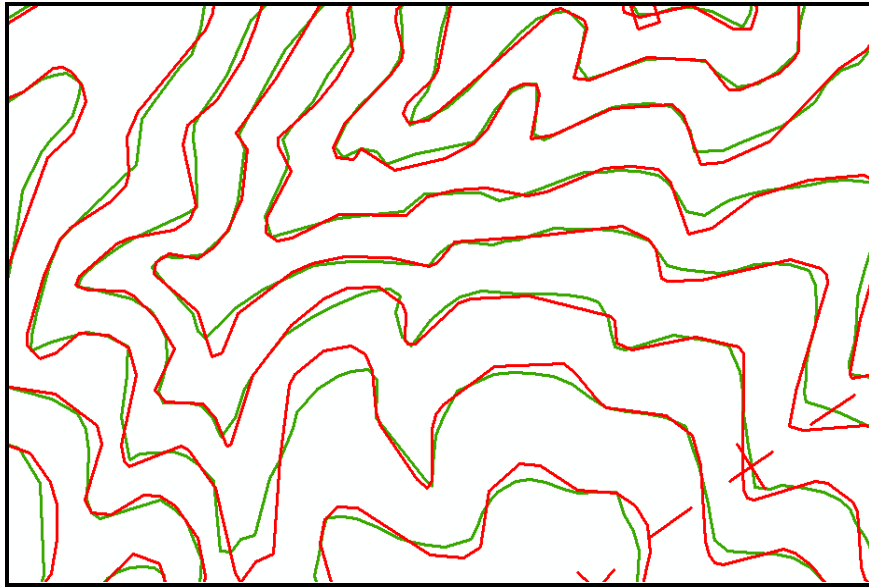


Figura 4.32 – Comparação das curvas generalizadas manualmente a vermelho e com o algoritmo TLopes a verde

É necessário efetuar mais experiências com outras folhas, outras escalas e outro tipo de morfologia do terreno para que possamos conhecer o valor a adotar para o caso da generalização da 25k para a 50k.

Embora as técnicas estatísticas fossem preferíveis para análise de resultados, nesta tese e porque se trata de dados geoespaciais foi também privilegiada a interpretação humana, uma vez que o que está em jogo é a representação cartográfica destes objetos, ou seja cumprir o objetivo da Base de Dados de médias escalas, que é no fundo atualizar e imprimir mapas nas escalas 25k e 50k. Compararam-se as curvas assim generalizadas com as correspondentes manuais, sendo as diferenças consideradas aceitáveis, segundo os padrões cartográficos de generalização manual, conforme mostrado nas figuras anteriores.

O algoritmo apresentado por TLopes apresenta algumas vantagens na sua utilização. Em primeiro lugar necessita de um único parâmetro para a sua utilização, sendo intuitiva a sua escolha, após uma análise geral da linha pelo operador. A curva de nível resultante mantém o essencial da sua forma atenuando os máximos e mínimos relativos. Os vértices resultantes não apresentam ângulos proeminentes, como os resultantes de algoritmos de simplificação com eliminação de pontos da linha. O resultado da sua utilização é previsível podendo o operador cartógrafo antever problemas e ter uma ideia muito clara do resultado antes da utilização do algoritmo.



# Capítulo 5

## Generalização cartográfica de linhas com recurso a Inteligência Artificial

---

### 5.1 Introdução

A generalização em qualquer NMA é efetuada, hoje em dia, de uma forma semiautomática, englobando duas ou mais séries cartográficas e abrangendo diversas escalas. Com esta abordagem obtemos uma simbiose homem/máquina, para efetuar uma tarefa onde, temos por um lado o conhecimento processual e cartográfico do lado do cartógrafo e os automatismos do lado da máquina. Por este motivo é necessário que o cartógrafo tenha muita experiência, seja um profundo conhecedor das regras cartográficas e no emprego dos algoritmos de generalização, antevendo possíveis resultados e problemas. Só detendo este nível de conhecimento e experiência se poderão obter bons níveis de execução e rentabilidade.

Pretende-se com este trabalho adquirir e utilizar, parte deste conhecimento, detido pelos cartógrafos e utilizá-lo na generalização de curvas de nível. Para tal é proposta uma metodologia, adaptada para o problema específico da generalização de curvas de nível da série 25k, para a série 50k. Contudo, esta metodologia poderá analogamente ser utilizada para outras escalas, ou mesmo até a outros temas da cartografia, utilizando outros algoritmos ou cadeias de algoritmos e respetivos parâmetros.

Este capítulo inicia-se com a apresentação da abordagem usada na generalização das curvas de nível. Seguindo-se a sua caracterização numérica e utilização da Inteligência Artificial, para o cálculo da tensão a aplicar no algoritmo de generalização. Incluindo

aqui a apresentação do treino e respetivos resultados. Nos pontos seguintes é efetuada a validação da metodologia e testes de sensibilidade ao modelo. Após a generalização das curvas de nível, pontos de cota e linhas de água é apresentada uma estratégia para a contextualização destes temas. Finalmente faz-se uma análise comparativa com outros métodos de generalização de curvas de nível, propostos por outros autores.

## 5.2 Estratégia de processamento

O método proposto para generalização de curvas de nível consiste na utilização de um algoritmo de generalização de linhas e na determinação automática dos parâmetros a utilizar no algoritmo em função das características da linha, com recurso a técnicas de inteligência artificial. Para o efeito são usadas duas bases de dados, uma base de dados geográficos com as curvas de nível e uma base de dados do conhecimento, que armazena as regras cartográficas utilizadas e as prioridades das entidades cartográficas, para efeitos de generalização (figura 5.1). O primeiro passo do método consiste na generalização modelo, no qual são seleccionadas as linhas, a partir da base de dados geográficos, a serem representadas na escala 50k. Num segundo passo é feita a caracterização numérica das curvas de nível, incluindo o cálculo da dimensão fractal, angularidade, média do comprimento dos segmentos, número de pontos. Os valores calculados são então registados nos atributos das linhas respetivas.

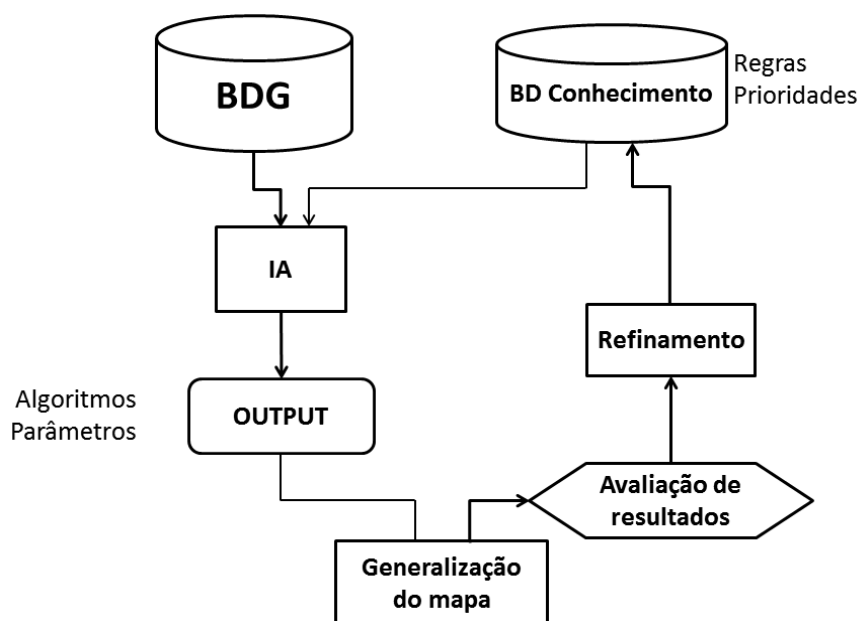


Figura 5.1 – Esquema geral da metodologia aplicada

Num terceiro passo, as curvas de nível são generalizadas com um algoritmo de generalização de linhas, com parametrização calculada previamente com os métodos de inteligência artificial e preservados na base de dados do conhecimento. O algoritmo de generalização de linhas deverá ser paramétrico, ou seja a simplificação ou suavização da linha é feita com base em um ou mais parâmetros. Sem perda de generalidade nesta tese foi usado o algoritmo TLopes (Lopes, 2006). Por último as curvas generalizadas são contextualizadas com os temas hidrografia e pontos cotados, com resolução automática de conflitos. O parâmetro a usar no algoritmo de generalização é determinado recorrendo aos métodos de IA, previamente treinados com uma amostra de curvas de nível. O método proposto será descrito em detalhe nas próximas secções.

### 5.3 Cálculo dos atributos das curvas de nível

Em primeiro lugar é efetuada a generalização modelo na qual são seleccionadas as curvas de nível com cotas múltiplas de 20 metros, valor da equidistância natural. São ainda eliminadas as curvas de nível que definem cabeços, “*curvas de nível fechadas*”, com comprimento inferior a 150 metros tal como os pontos cotados no seu interior, como especificado nas normas, excetuando-se aqui o caso dos Vértices Geodésicos. O tema altimetria compreende, para além das curvas de nível, os vértices geodésicos e os pontos cotados, conforme mostrado na figura 5.2. É necessário que o modelo conceptual tenha em conta estes objetos e o seu relacionamento.

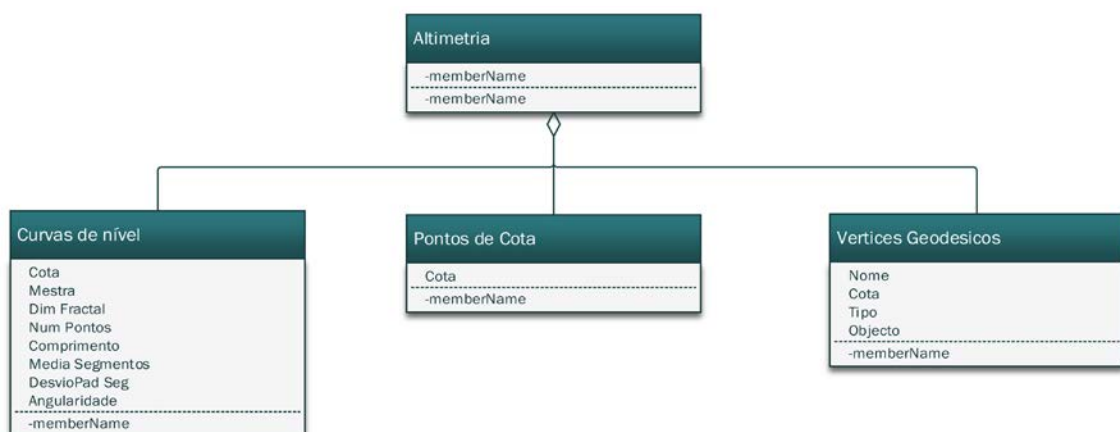


Figura 5.2 – Classes do tema altimetria

Após a criação das classes de objetos é necessário criar as respetivas instâncias, ou seja efetuar o carregamento da BDG e, calcular os atributos para cada uma, para tal recorreu-se à aplicação desenvolvida em ArcObjects, para o efeito.

O conjunto de atributos das linhas a calcular pela aplicação inclui:

- Dimensão Fractal;
- Número de vértices;
- Comprimento total da linha;
- Média do comprimento dos segmentos;
- Desvio padrão do comprimento dos segmentos;
- Angularidade;
- Altitude.

Para o cálculo da dimensão fractal utilizou-se o método da contagem das caixas. O método da *contagem das caixas* baseia-se na discretização de uma linha usando várias resoluções. Para calcular esta dimensão, coloca-se inicialmente uma grelha regular de dimensões  $s$  sobre a entidade e contam-se o número de caixas da grelha que são intersectadas pela entidade,  $N$ . O número  $N$  dependerá da escolha de  $s$ , consequentemente escreve-se  $N(s)$ . Muda-se  $s$  para tamanhos progressivamente menores e contam-se os números correspondentes  $N(s)$ . Em seguida faz-se um diagrama  $\log/\log$ , mais precisamente, um diagrama  $\log N(s)/\log(1/s)$ .

Para efeitos de programação é conveniente considerar uma sequência de grelhas onde o seu tamanho é reduzido sucessivamente por um factor de redução de  $1/2$ . Desta forma cada caixa da grelha é subdividida em quatro caixas cada uma de lado igual a metade do tamanho do lado da grelha anterior. Quando um fractal é calculado usando a contagem das caixas com estas grelhas chegamos a uma sequência de contagens  $N(2^{-k})$ , com  $k = 0, 1, 2, \dots$ . Convencionando que  $s = 2^0 = 1$  para a grelha maior, ver figura 5.3 e os passos respetivos para o cálculo da dimensão fractal.

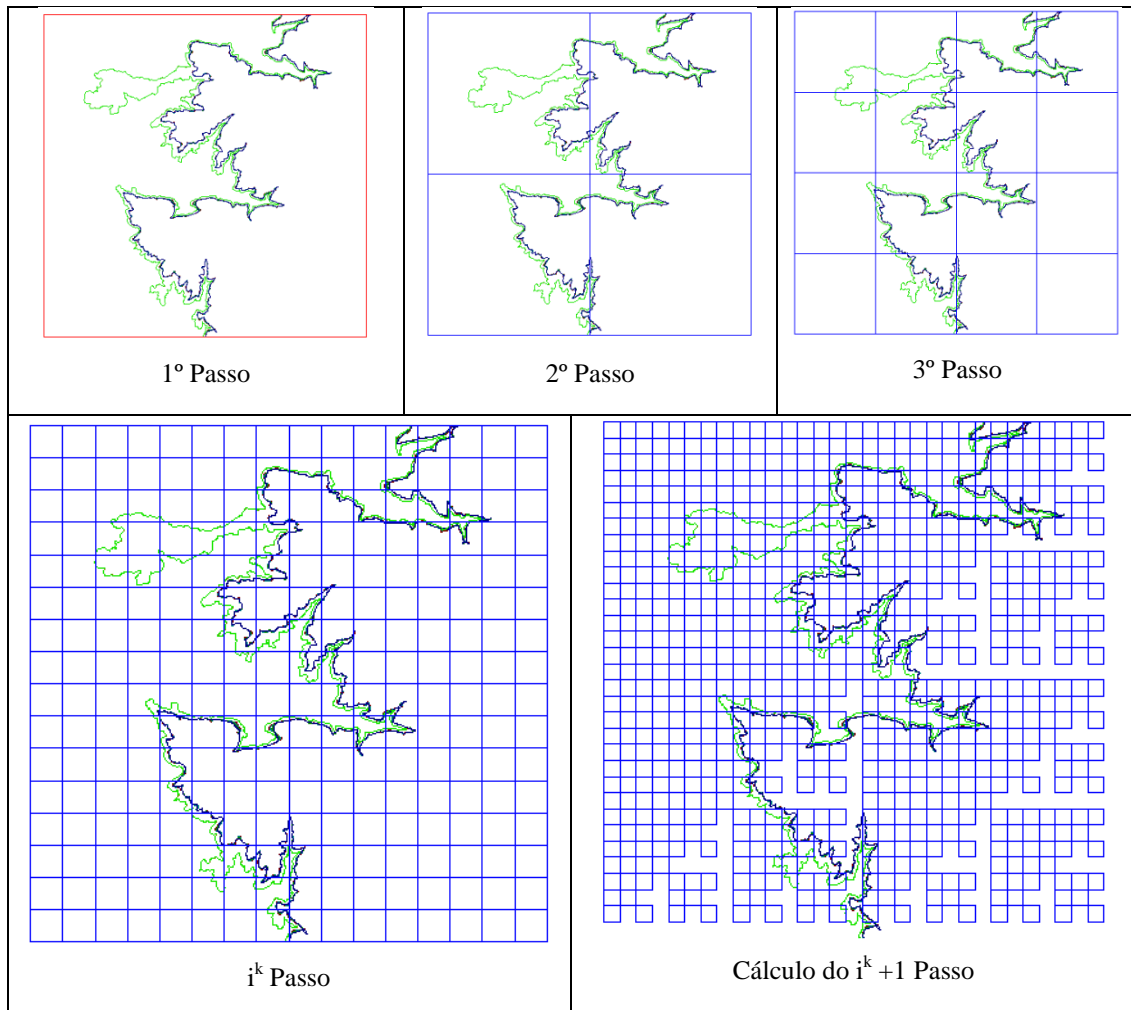


Figura 5.3 – Passos para o cálculo da Dimensão Fractal de uma curva de nível

O declive da linha recta ajustada no diagrama correspondente log/log é dado por (Peitgen, 1992) :

$$\frac{\log N(2^{-(k+1)}) - \log N(2^{-k})}{\log 2^{(k+1)} - \log 2^k} = \log_2 \frac{N(2^{-(k+1)})}{N(2^{-k})} \quad (5.1)$$

onde no termo da direita se usam logaritmos de base 2, enquanto o termo da esquerda resulta para qualquer base. O resultado é assim o logaritmo de base 2 do factor por que a contagem das caixas aumenta de uma grelha à seguinte. Este declive é uma estimativa para a dimensão fractal da “contagem das caixas”. Ou seja, se o número das caixas contadas aumentar por um factor de  $2^D$  quando o tamanho do lado da caixa é dividido ao meio, então a dimensão fractal é igual a  $D$ .

A dimensão fractal descreve a complexidade da linha, pois dá uma ideia de como a linha preenche o plano ou seja quanto mais complexa, maior a sua dimensão fractal, ver tabela 5.1.

Tabela 5.1 – Parâmetros característicos e sua descrição

Parâmetro	Descrição	Referência
Dimensão fractal (varia entre 1 e 2, para uma curva no plano)	Mandelbrot definiu <i>fractal</i> como - um sistema organizado para o qual a dimensão de Hausdorff-Besicovitch excede estritamente a dimensão topológica (número inteiro que caracteriza a geometria de um objeto euclidiano – por exemplo: zero para um ponto, um para uma linha, etc.).	(Bernhardt, 1992), (Jasinski, 1990).
Angularidade (varia entre 0 e $\pi/2$ )	Média dos ângulos entre segmentos consecutivos. (considera-se o menor ângulo, entre um segmento e o prolongamento do segmento anterior)	(Bernhardt, 1992), (McMaster, 1986), (Jasinski, 1990).
Número de vértices	Número total de vértices da linha.	(McMaster, 1986).
Comprimento da linha	Comprimento total da linha, soma do comprimento de todos os segmentos que a compõem.	(McMaster, 1986).
Média do comprimento dos segmentos	Média do comprimento da totalidade dos segmentos de uma linha.	(Jasinski, 1990).
Desvio padrão do comprimento dos segmentos	Desvio padrão da totalidade dos segmentos de uma linha.	(Jasinski, 1990).
Altitude das curvas de nível	Cota de cada curva de nível.	

## 5.4 Uso da IA na generalização de curvas de nível

Após a seleção das curvas de nível a representar na série 50k, é necessário generalizá-las e para isso recorreu-se ao algoritmo TLopes (Lopes, 2006), para o qual necessitamos definir para cada curva de nível o valor de tensão a aplicar, em função das características geométricas da curva. Esta dependência dos algoritmos, a utilizar na generalização de entidades, das respetivas características geométricas, foi observado por diversos autores como (Bernhardt, 1992; Balboa e López, 2008; Plazanet et al. 1998), embora utilizando outras técnicas e outros algoritmos. Pretende-se com a metodologia proposta, que a escolha do valor da tensão, passe a ser efetuada de forma inteligente como função das características da linha usando para o efeito algoritmos de inteligência artificial. São usados três métodos de IA na estimativa do valor da tensão, os seus

*outputs* são depois comparados e analisados por um agente, num leilão, onde é escolhido o “*melhor*” valor para o parâmetro tensão, a utilizar na generalização das curvas de nível.

Para o cálculo da tensão a usar para cada linha, usou-se uma Rede Neuronal, uma Árvore de Classificação e Regressão e uma Árvore de Decisão, recorrendo ao algoritmo C5, de modo a prever o parâmetro tensão, usado no algoritmo de generalização de curvas de nível. Para qualquer destes métodos de IA, utilizaram-se os atributos da linha, já mencionados, como *input*, para obter o valor da tensão, como *output* (fig. 5.4).

Como descrito anteriormente, é necessário treinar a rede neuronal e construir as árvores de decisão e classificação e regressão, com dados representativos do conjunto que se pretende generalizar. Para o treino da rede neuronal e construção das árvores de decisão e de classificação e regressão, é necessário previamente obter os atributos caracterizadores das curvas de nível, assim como os valores da tensão seleccionados por um operador experiente, fig. 5.5a) e b). Após o treino o classificador determina o valor da tensão a utilizar na generalização das diferentes curvas de nível.

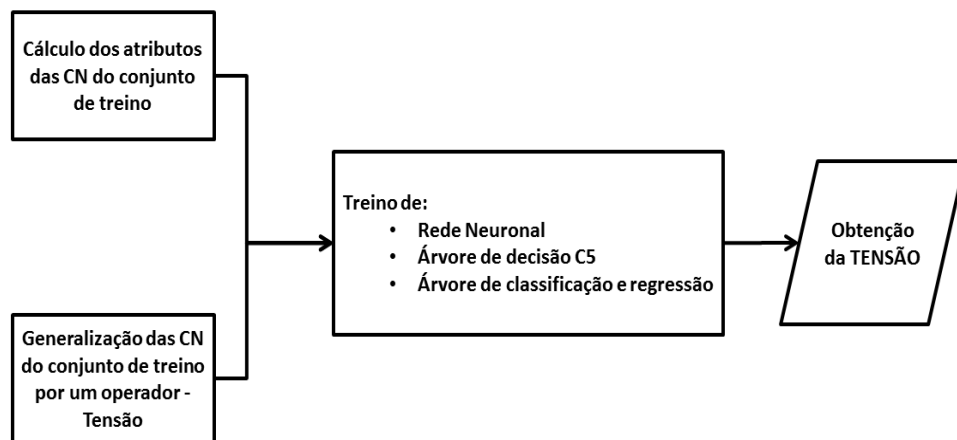


Figura 5.4 – Fases do treino

Após a aprendizagem e para os novos casos, os atributos característicos das curvas de nível são conectados à rede neuronal, à árvore de classificação e regressão e árvore de decisão que, com os parâmetros previamente refinados durante o processo de aprendizagem, produzem como saída um valor de tensão a utilizar no algoritmo de generalização, ver figura 5.5 d1) d2) e d3), (*NN-Rede Neuronal, C&R-Árvore de Classificação e Regressão, C5-Árvore de Decisão recorrendo ao algoritmo C5*).

Neste caso foi utilizada uma Rede Neuronal multicamada. Foram testadas várias redes para decidir sobre o número de camadas escondidas, após estas experiências concluiu-se que para este processo e com estes dados, os melhores resultados foram alcançados com 3 camadas intermédias escondidas. A camada de entrada integra 7 unidades, uma por cada atributo numérico caracterizador da curva de nível. Na camada de saída é apresentado o valor do parâmetro (tensão) a utilizar no algoritmo, ver figura 5.5 d<sub>1</sub>). Para função de ativação utilizou-se uma sigmoide de 0.5.

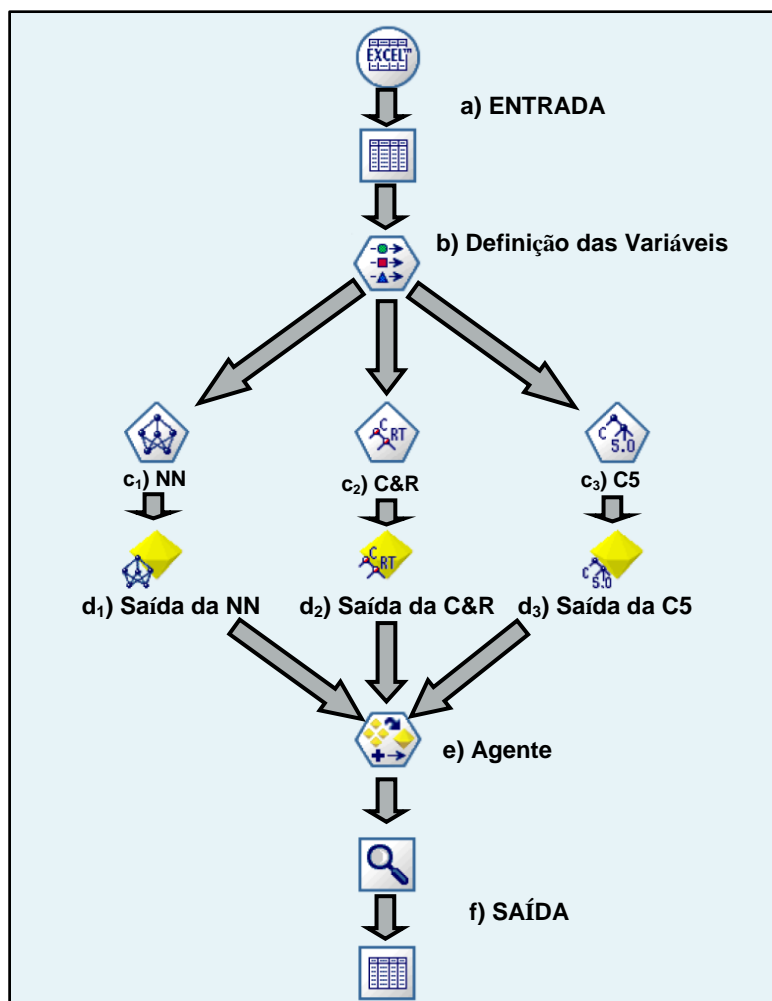


Figura 5.5 – Projeto Geral de atividades de Inteligência Artificial

Os resultados destes três métodos são combinados num nó, onde um agente é responsável por um “leilão”, de modo a que o valor de tensão selecionado seja o “melhor” dos resultantes dos três métodos. Este agente combina dois ou mais *Outputs* para obter predições mais precisas, que as dadas por cada método individualmente. Combinando predições de múltiplos métodos, as limitações destes métodos individuais podem ser corrigidas ou atenuadas, tendo como resultado final uma precisão superior,



ver figura 5.5e). Na combinação destes métodos obtém-se um resultado, pelo menos tão bom como o melhor dos métodos individuais, sendo frequentemente melhor. Baseado nos atributos das curvas de nível, foi possível prever o valor da tensão a usar no algoritmo de generalização, ver figura 5.5f), libertando o cartógrafo desta tarefa.

É necessário que se tenha a noção que embora o valor seja ‘o melhor’, tal não dispensa a necessidade de avaliação efetuada por um cartógrafo. Pois sempre que se utiliza um algoritmo para generalizar, poderemos ter um bom resultado em grande parte da linha e, ter outra parte da linha em que o resultado alcançado não seja aceitável.

Depois do treino e validação é possível tirar partido, da rede neuronal treinada para estes casos e, das árvores de decisão e de classificação e regressão de forma a generalizar as curvas de nível das restantes folhas da série 25k, para serem representadas na série 50k, ver figura 5.6.

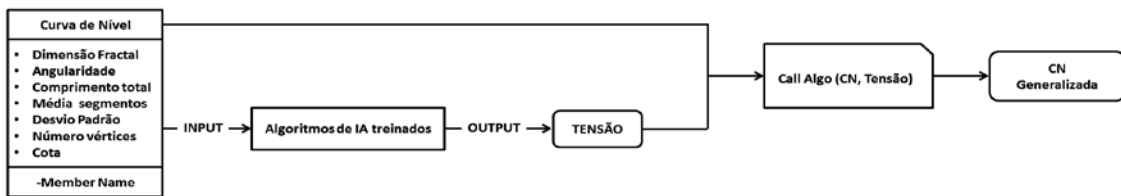


Figura 5.6 – Generalização das Curvas de Nível

Os atributos caracterizadores das curvas de nível são lidos a partir da BDG e utilizados como entrada dos métodos de IA. Com base nestes atributos é calculado o valor da tensão a aplicar à linha. O algoritmo de generalização é depois chamado para efetuar a generalização da linha com o valor da tensão obtido previamente, conforme figura 5.6.

## 5.5 Treino dos classificadores

Efetuaram-se duas experiências com mapas da série 25k do IGeoE (*Instituto Geográfico do Exército*), com o objetivo de generalizar as curvas de nível destes mapas para a escala 50k. Os resultados foram posteriormente comparados com a generalização manual. Foram usadas as curvas de nível da folha N° 309 da série 25k, num total de 249 linhas, para construção do modelo e treino dos métodos de IA. A folha N° 50 foi utilizada para a validação do método, utilizaram-se ainda 4 folhas para os testes de sensibilidade, generalizadas por 4 operadores e finalmente a folha N° 279 foi generalizada também por 4 operadores, ver figura 5.7.

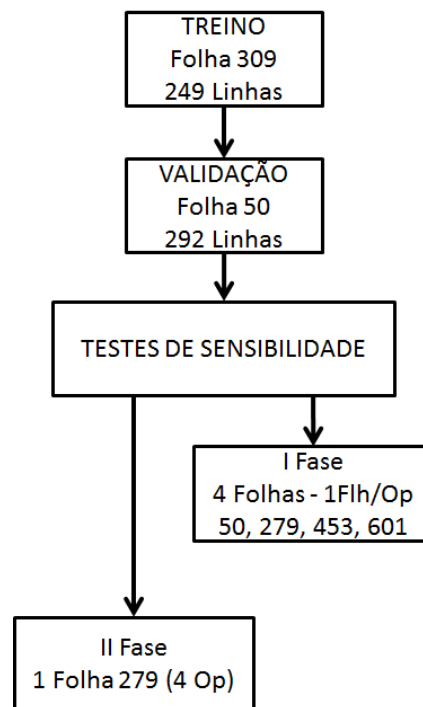


Figura 5.7 – Esquema geral das atividades de treino, validação e testes

O sistema necessita de ser treinado com dados reais. Para isso um operador experiente seleccionou a tensão óptima, a usar no algoritmo de generalização de linhas, a ser aplicada para cada curva de nível, de modo a ser generalizada da escala 25k para a escala 50k. Foram processados dois conjuntos de curvas de nível: o primeiro para treinar os classificadores e o segundo para avaliar a qualidade do processo de generalização.

Tabela 5.2 – Estatística da caracterização de linhas

Parâmetro	Min	Max	Mediana	Média
Dimensão Fractal	1.049	1.577	1.361	1.351
Número de vértices	10	11339	77	475
Comprimento da linha (m)	50.7	107383.5	700.8	4575.1
Angularidade (grau)	5.2	34.0	9.6	10.0
Média do comprimento dos segmentos (m)	4.9	13.4	8.8	9.0
Desvio padrão dos segmentos (m)	4.1	698.1	48.1	86.9
Altitude das curvas de nível (m)	100	500	240	265

As curvas de nível utilizadas no processo de aprendizagem englobam um grande conjunto de formas, como se pode ver na tabela 5.2, a sua dimensão fractal varia de 1.049 a 1.577, o que significa que a sua sinuosidade vai desde o quase plano até ao muito sinuoso, o número de vértices varia entre 10 e 11339 conjugado com o

comprimento da linha, média do comprimento dos seus segmentos e respetivo desvio padrão é indicativo da sua complexidade, resultando num conjunto de linhas abrangente e demonstrativo como podemos ver na figura 5.8.

Um operador procedeu à generalização das curvas de nível usando para o efeito o algoritmo TLopes. Deste modo foi atribuído a cada curva de nível um valor para a tensão a usar no algoritmo. Para cada curva e implicitamente para cada conjunto de características de uma linha, foi estabelecida a correspondência com um valor de tensão. Este conjunto de valores é usado para treinar o classificador. Neste caso foi usado para treinar 3 classificadores: C&R, C5 e Rede neuronal. Na tabela 5.3 é apresentada o peso normalizado de cada uma das características da linha no processo de estimação da tensão a aplicar à linha, determinado por cada classificador.

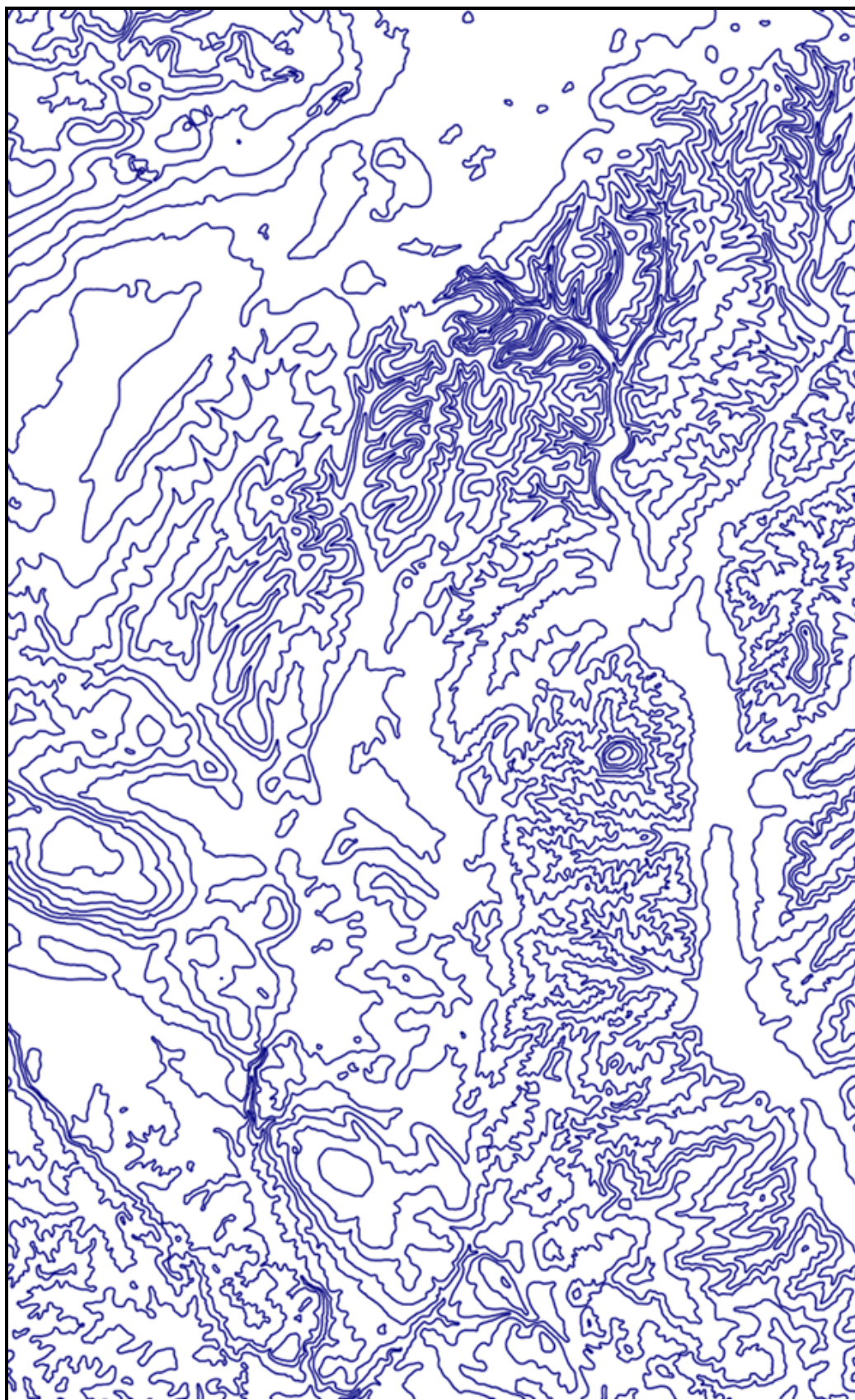
Tabela 5.3 – Importância das variáveis no processo de classificação

<b>Variáveis</b>	<b>C&amp;R</b>	<b>C5</b>	<b>Rede Neuronal</b>
Dimensão Fractal	0.642	0.697	0.314
Número de vértices	0.000	0.012	0.104
Comprimento da linha	0.008	0.000	0.117
Angularidade	0.083	0.115	0.109
Média do comprimento dos segmentos	0.243	0.041	0.198
Desvio padrão dos segmentos	0.008	0.093	0.069
Altitude das curvas de nível	0.017	0.042	0.090

Analisando a tabela 5.3 podemos constatar que para qualquer dos métodos, a dimensão fractal é aquela que apresenta a maior importância para a classificação, em contraste com o número de vértices e o comprimento da linha que não influenciam grandemente esta classificação.

Os parâmetros estimados pelos classificadores e selecionados pelo agente, foram aplicados à folha 309 da série 25k, com os quais se generalizaram as respetivas curvas de nível. A figura 5.8 representa as curvas de nível originais da folha 309, na escala 25k, utilizada na aprendizagem dos métodos de IA. Conforme podemos ver estas apresentam uma grande variedade de formas.

Figura 5.8 – Curvas de nível, da folha 309 à escala 25k



Na figura 5.9 representam-se extratos desta folha, onde podemos ver a comparação entre as curvas de nível originais na escala 25k e generalizadas na escala 50k.

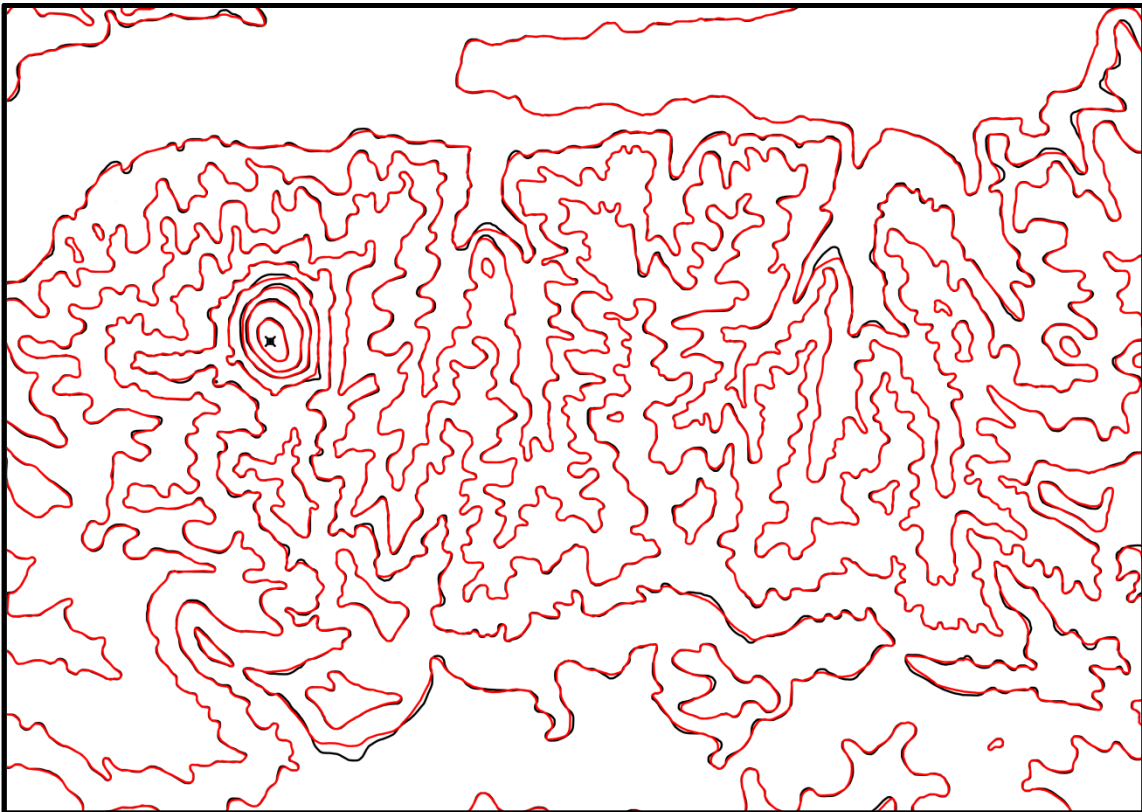


Figura 5.9a – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais

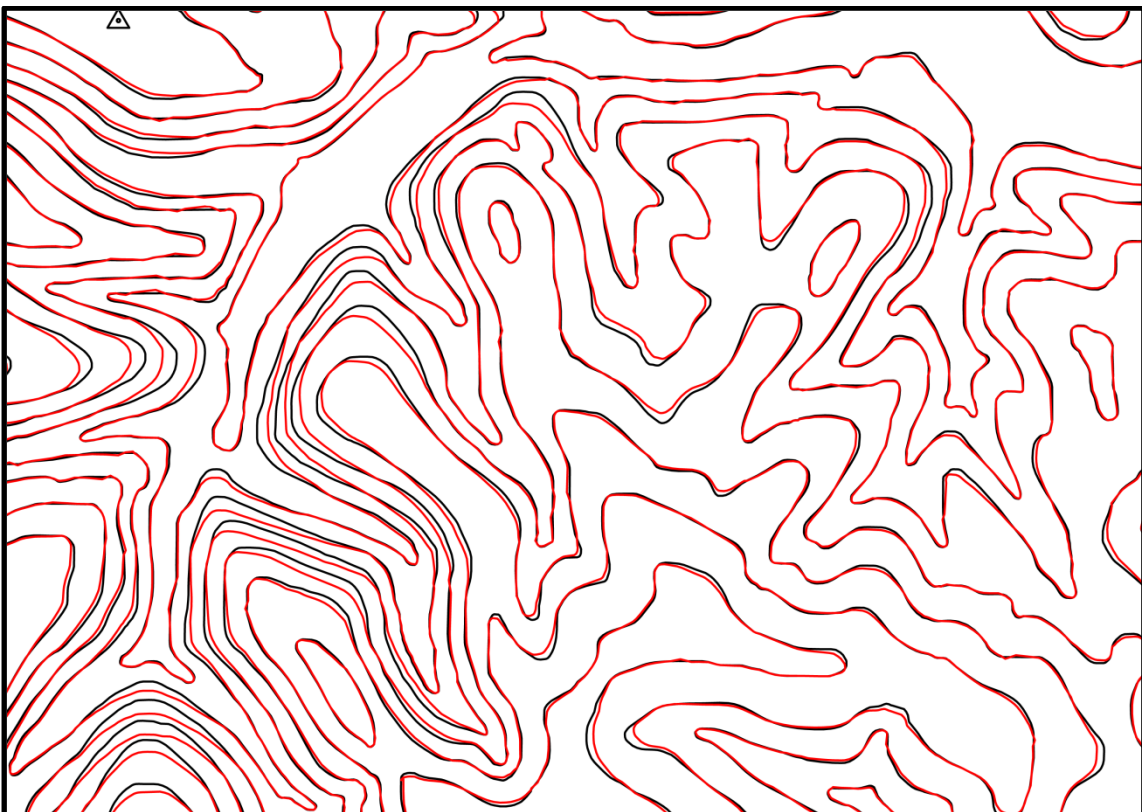
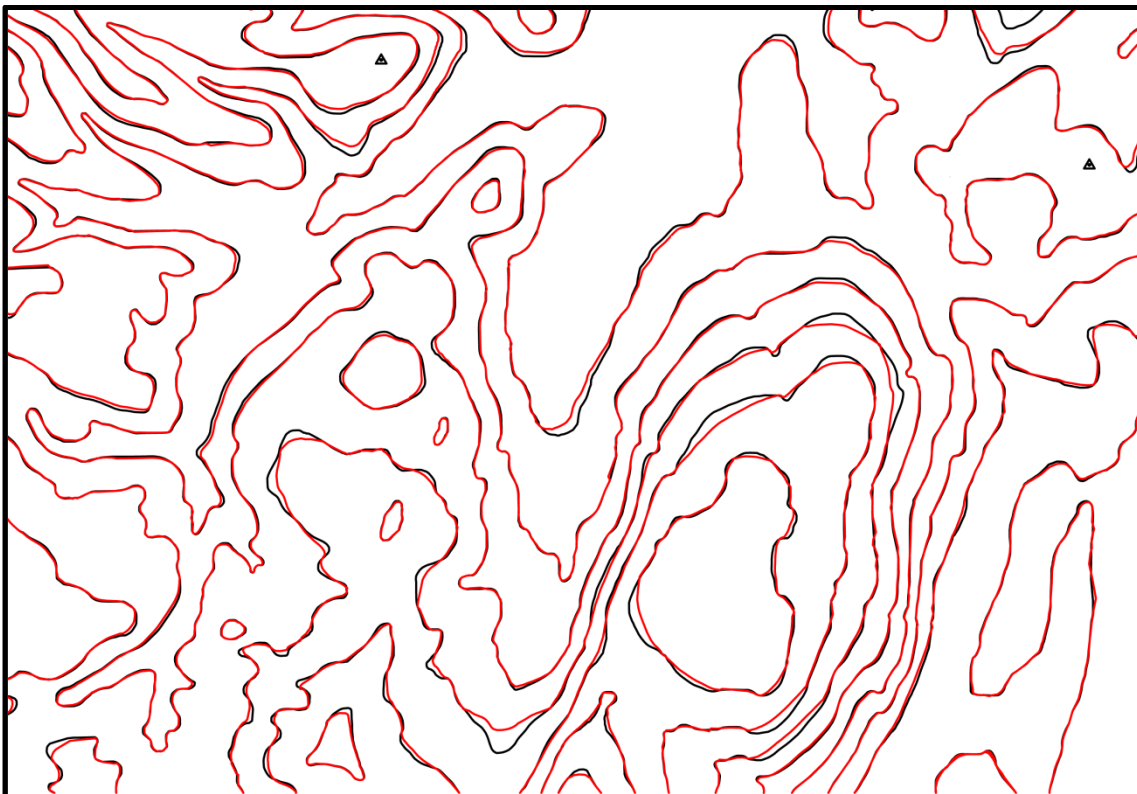
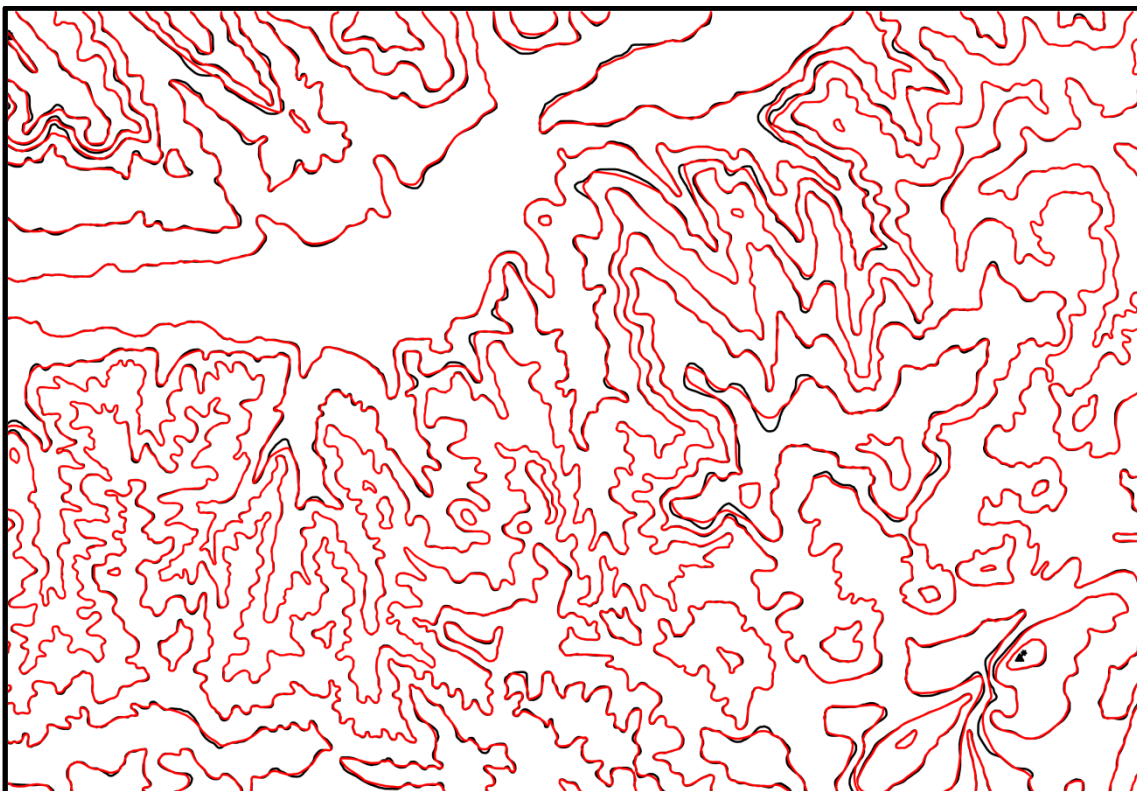


Figura 5.9b – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais





**Figura 5.9c** – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais



**Figura 5.9d** – Extrato das curvas de nível, da folha 309 à escala 25k, a vermelho curvas generalizadas, a preto curvas originais

## 5.6 Validação

A avaliação da qualidade foi efetuada a três níveis: comparação visual das curvas de nível generalizadas com o método proposto nesta tese e as curvas generalizadas manualmente, comparação dos valores de tensão gerados pela técnica e os valores de tensão sugeridos por um operador e por fim controlo de qualidade efetuado na SCQ (Secção de Controlo de Qualidade) do IGeoE.

A qualidade do resultado, foi testada por comparação entre as curvas resultantes do método proposto e as curvas resultantes da generalização efetuada por um cartógrafo experiente. Para validação do processo foram usadas 292 curvas de nível da folha N°50 com base nas quais se compararam os valores de tensão estimados e os valores de tensão escolhidos por um operador. As colunas representam a tensão indicada pelo cartógrafo e as linhas correspondem à tensão estimada pelo algoritmo. Analisou-se separadamente a Rede Neuronal (tabela 5.4), a árvore de decisão com o algoritmo C5 (tabela 5.5), a árvore de classificação e regressão (tabela 5.6) e após a utilização do agente (tabela 5.7).

Tabela 5.4 - Matriz de confusão para a Rede Neuronal.

**Operador/Rede Neuronal**

	0	1	2	3	4	5	6	7	8	9	10	Total
0												
1		13	53									66
2			61		1	1						63
3			5	2								7
4			10			1						11
5			15	1	2	120	3	1	1			143
6							2					2
7												
8												
9												
10												
Total		13	144	3	3	122	3	3	1			292

Pc – Percentagem de concordância 67%

Tabela 5.5 - Matriz de confusão para a Árvore C5.

**Operador/Árvore de decisão C5**

	0	1	2	3	4	5	6	7	8	9	10	Total
0												
1			7									7
2			13	128	2	1	36					180
3												
4				2		1	19					22
5				7	1	1	54	2	1			66
6							3	1				4
7							10		2	1		13
8												
9												
10												
Total		13	144	3	3	122	3	3	1			292

Pc – Percentagem de concordância 63%

Tabela 5.6 - Matriz de confusão para a Árvore C&R

**Operador/Árvore C&R**

	0	1	2	3	4	5	6	7	8	9	10	Total
0												
1			7									7
2			13	129	2	1	35					180
3												
4				2		1	19					22
5				6	1	1	55	2	1			66
6							3	1				4
7							10		2	1		13
8												
9												
10												
Total		13	144	3	3	122	3	3	1			292

Pc – Percentagem de concordância 64%

Tabela 5.7 - Matriz de confusão depois do Agente.

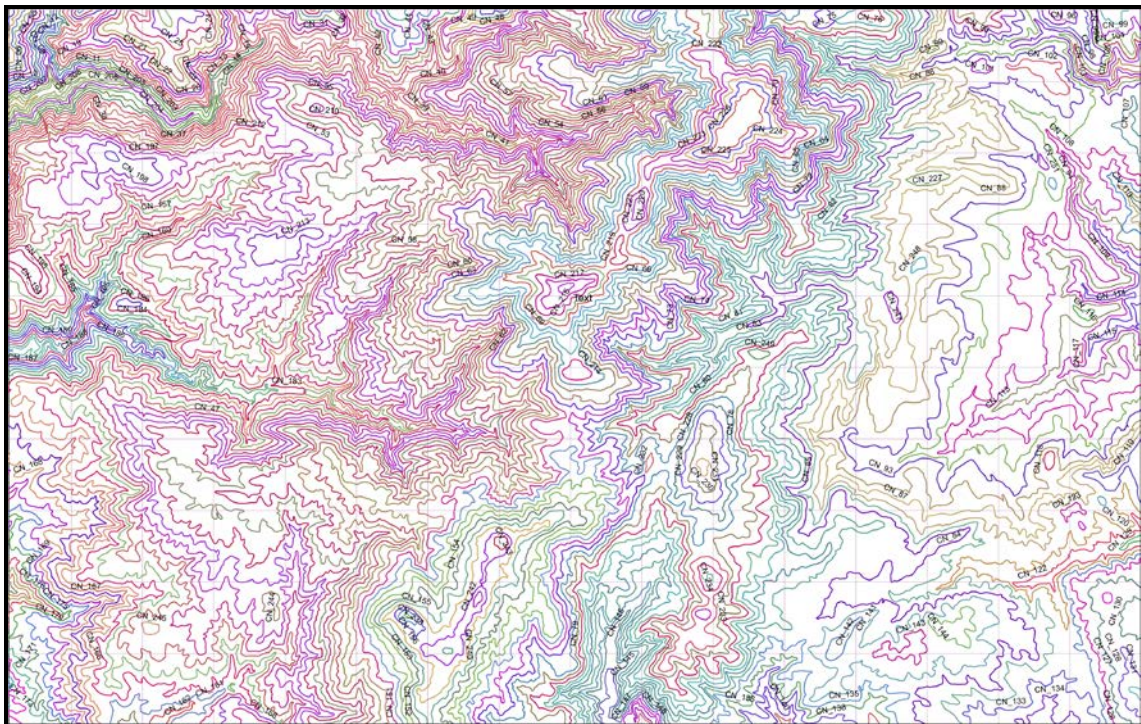
**Operador/Agente**

	0	1	2	3	4	5	6	7	8	9	10	Total
0												
1		5	1									6
2		8	126	2		15						151
3					1							1
4				6		1	5					12
5				11	1	1	102	3	1	1		120
6												
7								2				2
8												
9												
10												
Total		13	144	3	3	122	3	3	1			292

Pc – Percentagem de concordância 81%

Na análise destas matrizes podemos verificar que os valores se distribuem em torno da sua diagonal principal, significando haver concordância relativa entre as tensões propostas pelo operador e as calculadas pelos métodos de IA. As maiores tensões não são usadas, nem pelo operador nem pelos métodos de IA, pelo que para o rácio destas escalas a deformação das linhas não deve ser muito elevada. Com a utilização do agente obtemos um ganho considerável na predição do valor da tensão a usar, quando combinamos os três métodos.

A percentagem de concordância aumenta significativamente com a utilização do agente, obtendo um valor de 81%. Considerando que a diferença na deformação das linhas não é excessivo para valores de tensão próximos, então o número de curvas de nível aceitáveis poderá ser superior. No entanto em algumas linhas o desvio poderá ser apenas parcialmente aceitável, pelo que é necessário a supervisão de um operador cartógrafo. Na figura 5.10 apresentam-se as curvas de nível da folha N° 50 generalizadas com a metodologia proposta, utilizada para validação. Como se pode constatar grande parte das curvas são admissíveis necessitando de pouco trabalho de operador.



**Figura 5.10 – Resultados da metodologia aplicada a curvas de nível da folha 50**

A figura 5.11 apresenta extratos das folhas cujas curvas de nível foram generalizadas com a metodologia aqui apresentada. Podemos verificar que as curvas de nível são adequadas para a representação na escala 50k.



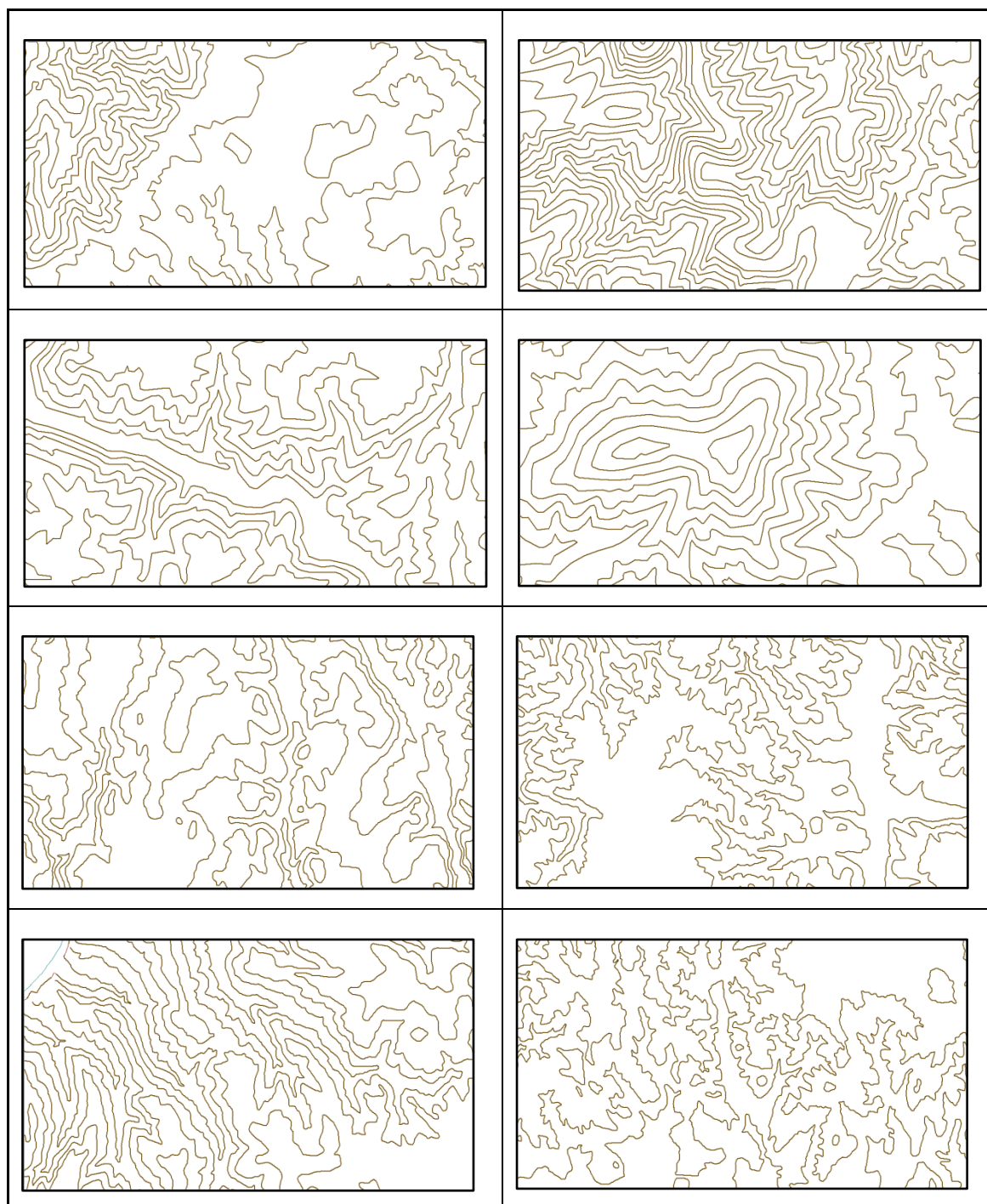


Figura 5.11 – Resultados da metodologia aplicada a curvas de nível, extratos das folhas utilizadas

### 5.6.1 Validação qualitativa

Esta validação foi efetuada na Secção de Controlo de Qualidade do IGeoE, onde as folhas, generalizadas com a metodologia apresentada, foram submetidas ao processo de controlo de qualidade. A validação quantitativa é notoriamente insuficiente para estes

casos de cartografia, pois mesmo apresentando valores da tensão a utilizar, concordantes entre o método aqui apresentado e os valores propostos pelos operadores, é necessário verificar se as regras cartográficas não foram comprometidas, assim como a topologia ou mesmo a legibilidade do produto final. Todos estes aspetos devem ser tidos em conta.

Apenas com uma visualização da representação dos dados por um operador cartógrafo, esta avaliação é possível, não podemos dispensar esta avaliação qualitativa sob o risco de obter um mapa não legível ou em condições de utilização muito insipientes. Foi ainda pedido aos operadores da SCQ que enunciassem três pontos fortes e três pontos fracos desta metodologia e respetivos resultados, o resultado é resumido nos pontos seguintes.

**Pontos fracos e a melhorar:**

- A tensão não deveria ser aplicada igualmente a toda a linha, mas deveria ser aplicada diferentemente a determinados troços da linha, conforme a rugosidade da linha.
- O algoritmo apresenta comportamentos diferentes para locais aparentemente iguais das linhas.
- As linhas de água não estão centradas com as inflexões das curvas de nível.
- Deve permitir a eliminação de um maior número de vértices.
- Em alguns casos a tensão máxima deforma a linha não ficando reconhecível.
- Pormenores assinalados pela secção de controlo de qualidade, curvas de nível muito próximas das linhas de água, ver figura 5.12.
- Algumas inflexões apresentam um aspeto muito anguloso, ver figura 5.12.
- Não tem em conta alguns pormenores cartográficos que pode levar a erros.
- Na generalização da 25k para a 50k algumas inflexões deveriam simplesmente deixar de existir, visto não serem representativas à nova escala, ver figura 5.12.
- Verifica-se que as curvas de nível ficam com muitos vértices quando a tensão é maior.
- A tensão máxima em alguns casos pode levar a uma alteração do traçado da linha de água ou retirar importância à mesma.

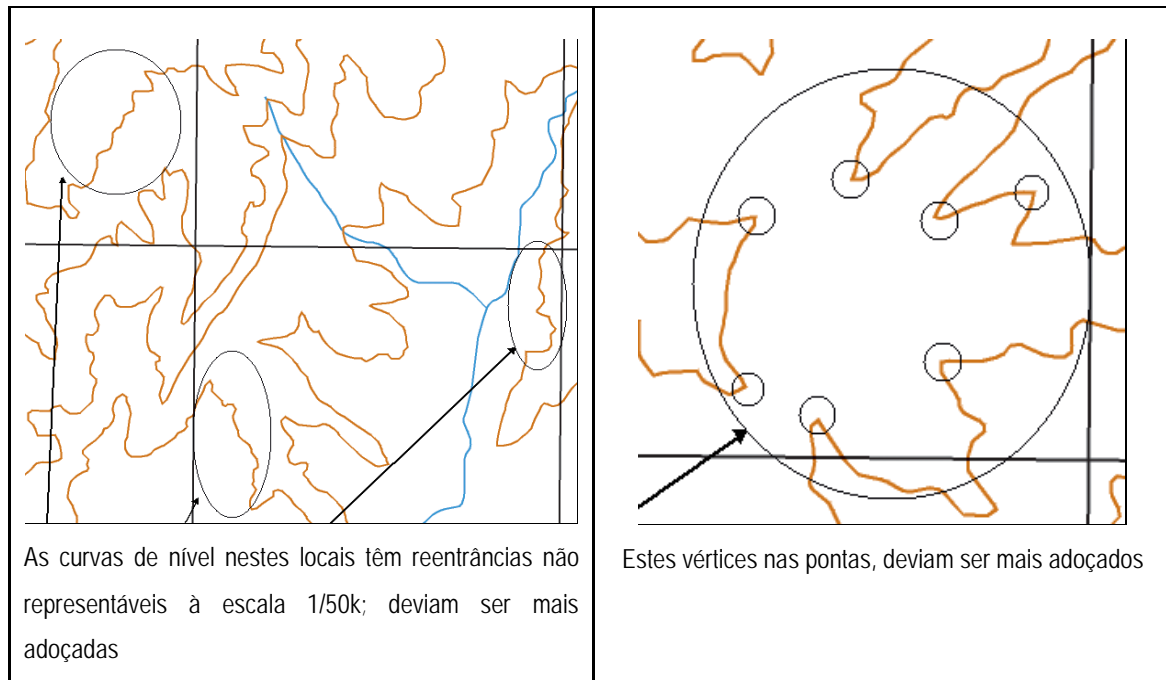


Figura 5.12 – Erros apontados pelos operadores da Secção de Controlo de Qualidade

#### Pontos fortes:

- Altimetria harmoniosa de acordo com a escala.
- Fácil e prático para o operador, com poucas intervenções no processo.
- Facilidade de alteração do valor da tensão a aplicar, e modificar pelo operador caso assim o entenda.
- Quando a curva de nível original tiver muitas e pequenas variações pode-se aplicar a tensão máxima, pois os vértices não são visíveis.
- O resultado final requer poucas alterações efetuadas pelo operador.
- Igualdade de critérios para as folhas adjacentes, o que não se verifica quando as folhas são generalizadas por diferentes operadores cartógrafos.
- Não cria problemas topológicos, nomeadamente com os pontos de cota e vértices geodésicos.

Os mapas resultantes foram comparados com os manualmente generalizados. Estes resultados revelam que apesar de haver algumas diferenças entre as curvas de nível, o resultado é admissível. O trabalho final de operador neste tema é reduzido, havendo um ganho efetivo em meios e tempo.

### **5.6.2 Testes de sensibilidade do modelo**

Para testar o modelo de uma forma mais robusta, efetuaram-se generalizações da altimetria de folhas de zonas morfológicamente diferentes, folha N°50 do norte do país, folha N°279 e N°453 do centro e folha N°601 do sul. Após a generalização as folhas foram submetidas a testes de validação qualitativos, efetuados por diversos operadores cartógrafos. O tempo de operador gasto em média por folha em generalização manual de curvas de nível, é de 35 horas, com esta metodologia pode ser reduzido para cerca de 3 a 4 horas. Apesar de haver a avaliação e correções finais a realizar pelo operador. Na figura 5.13 são apresentadas as curvas de nível da carta 50k generalizadas com os parâmetros preditos pelas técnicas apresentadas.

A sensibilidade do modelo foi testada em duas fases. Numa primeira fase, utilizaram-se 4 folhas da série 25k e 4 operadores cartógrafos utilizaram o algoritmo apresentado, generalizando as curvas de nível de uma folha diferente da utilizada na aprendizagem dos métodos de IA. O resultado destas generalizações foi comparado com as correspondentes folhas da série M782, escala 50k. Foi ainda efetuada uma validação qualitativa, ver ponto 5.6.1.

Numa segunda fase deste teste além das folhas anteriores, quatro operadores cartógrafos generalizaram, independentemente, a folha N° 279, num total de 255 curvas de nível. Na figura 5.14 são apresentados os histogramas do valor da tensão selecionado por cada operador. Da análise dos histogramas da figura 5.14, podemos ver que os operadores não concordam entre si, no valor da tensão a utilizar, para alguns casos. No entanto, para o primeiro conjunto de linhas, embora os operadores utilizem diferentes valores de tensão, concordam entre si na aplicação do mesmo valor de tensão neste primeiro grupo de linhas. Nota-se ainda que o operador 4 utiliza a tensão máxima numa série de linhas, preferindo utilizar as tensões mais elevadas, havendo uma maior discordância entre este operador e os restantes, o operador 3 utiliza o valor de tensão 0 em algumas linhas, não sendo muito utilizado pelos restantes. Contudo o operador 1 utiliza o valor de tensão 0 nas primeiras curvas, não o voltando a utilizar, nas restantes. O operador 2 utiliza basicamente 2 valores de tensão, a tensão 4 e a tensão 8, para toda a folha, com exceção de algumas linhas. Os operadores 1 e 2 apresentam também algumas discordâncias, mas menores, concordando numa grande parte do conjunto de linhas.



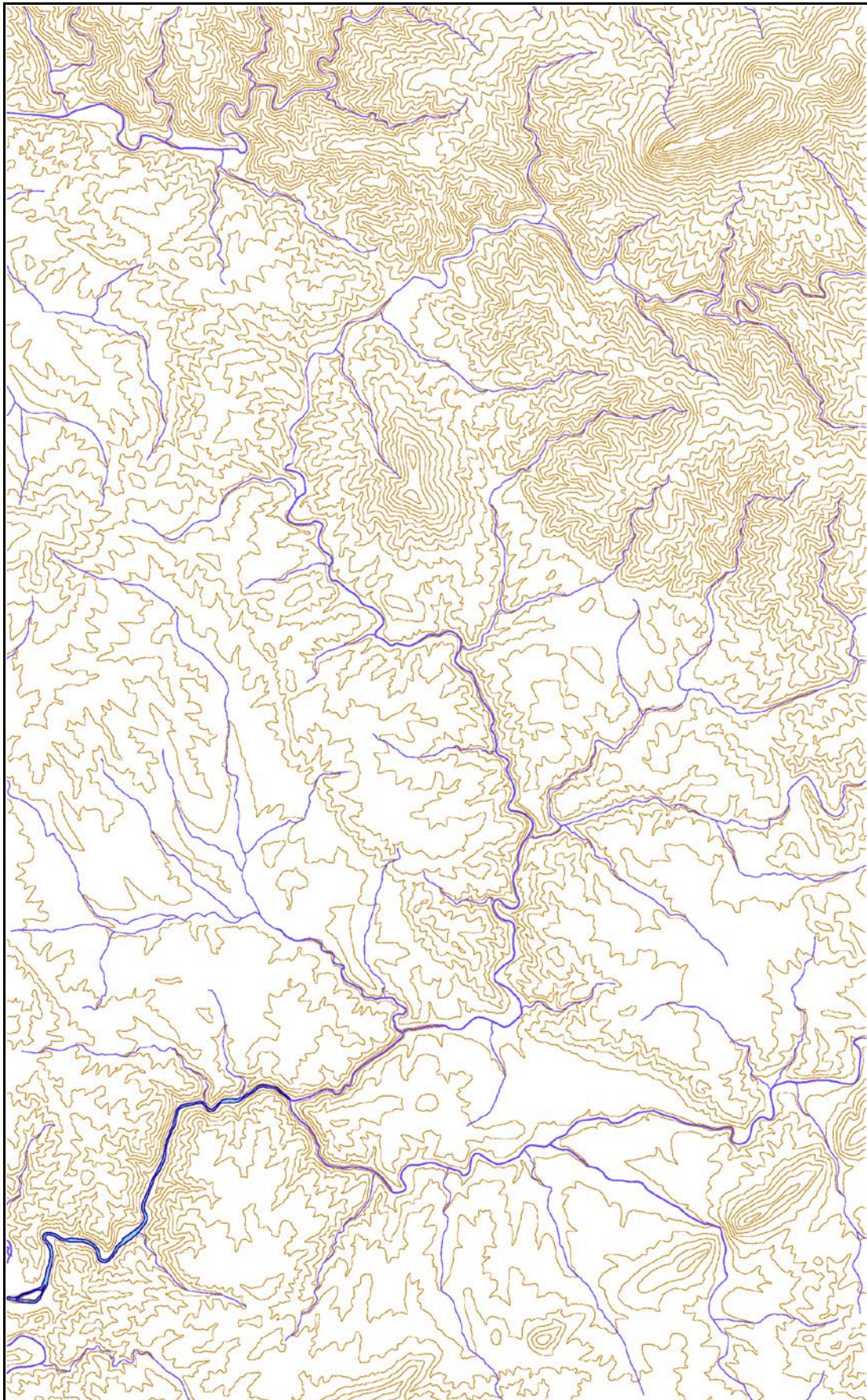


Figura 5.13 – Curvas de nível e linhas de água generalizadas para a escala 1/50 000

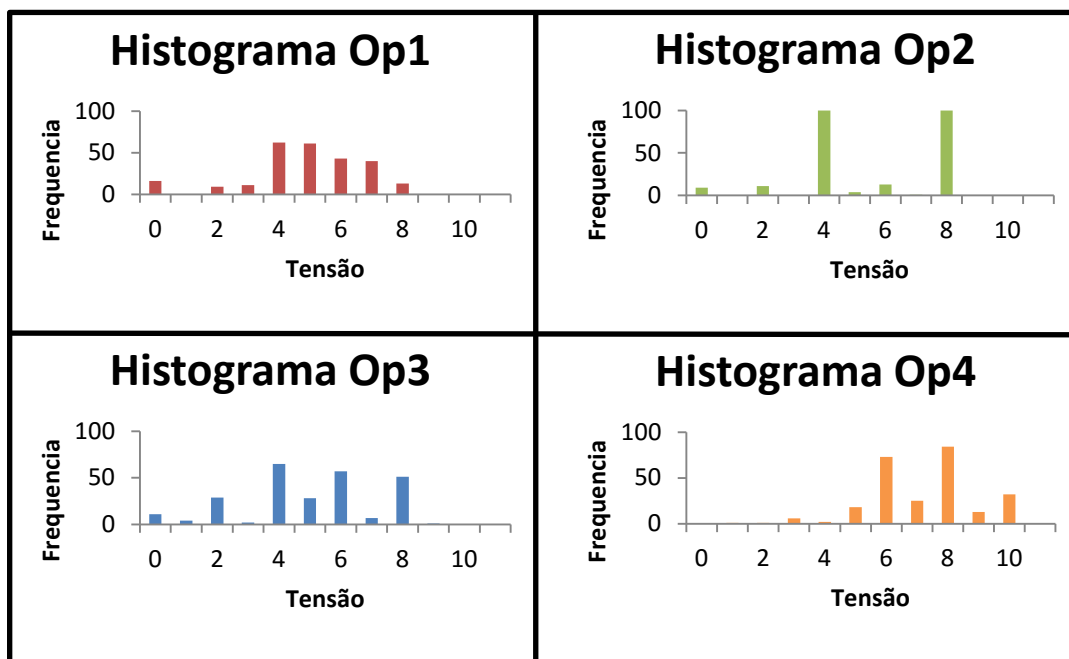


Figura 5.14 – Histogramas com os valores da tensão utilizada pelos quatro operadores (Op1, Op2, Op3 e Op4) para uma mesma folha (N. 279)

Estas constatações estão de acordo com o que alguns autores já tinham observado, ver Capítulo 3 desta tese, por um lado existem várias soluções possíveis e admissíveis, por outro o conhecimento dos cartógrafos não apresenta um padrão, ou seja não existe consenso. Este facto torna mais complexa a utilização e treino dos classificadores em generalização. Por este motivo optou-se por voltar a treinar a rede Neuronal e construir as Árvores de Decisão e Classificação e Regressão com os valores de tensão médios destes operadores e nalguns casos justificados, os valores específicos da escolha de um ou outro operador cartógrafo, como por exemplo quando aplicado um valor de tensão baixo pelo facto da curva de nível ter um comprimento e sinuosidade pequenos.

Estas constatações indicam-nos que devemos refinar os métodos de IA, com o aparecimento de novos casos que serão acrescentados às tabelas de treino, contribuindo desta forma para que este número de casos discrepantes, vá sendo cada vez menor e a sua eficácia vá aumentando. Caso a aplicação o permita, podem treinar-se os classificadores acrescentando ao histórico, da aprendizagem já efetuada, os novos casos, poupando assim recursos e tempo de processamento.

Após a generalização da folha N° 279, recorrendo à metodologia apresentada, compararam-se as tensões utilizadas pelos operadores e respetivos métodos de



Inteligência Artificial, cujos resultados se resumem nas matrizes de confusão das tabelas 5.8, 5.9, 5.10 e 5.11.

Tabela 5.8 - Matriz de confusão para a Rede Neuronal.

**Operadores/Rede Neuronal**

	0	1	2	3	4	5	6	7	8	9	10	Oper
0												
1		9		2	1							12
2		4	33	7	1							45
3		4	9	45	5		1					64
4			1	3	33	5	1					43
5				3	6	37	2	1				49
6				1	2	3	19		1			26
7						1	1	9				11
8								1	4			5
9												
10												
RN		17	43	61	48	46	24	11	5			255

Pc – Percentagem de concordância 74%

Tabela 5.9 - Matriz de confusão para a Árvore C5.

**Operadores/Árvore de decisão C5**

	0	1	2	3	4	5	6	7	8	9	10	Oper
0												
1		10		2	1							13
2		6	29	2	1							38
3		1	10	50	3	1	1					66
4			3	6	35	3	3					50
5			1	1	7	40	10	1	1			61
6					1	1	10	5	2			19
7						1		5				6
8									2			2
9												
10												
C5		17	43	61	48	46	24	11	5			255

Pc – Percentagem de concordância 70%

Tabela 5.10 - Matriz de confusão para a Árvore C&R

**Operadores/Árvore C&R**

	0	1	2	3	4	5	6	7	8	9	10	Oper
0												
1		8	1									9
2		7	31	10	2	1						51
3		2	8	43	2	1	1					57
4			3	4	39	10	4					60
5				4	4	32	9	1	1			51
6					1	1	10	7	4			23
7						1		3				4
8												
9												
10												
CART		17	43	61	48	46	24	11	5			255

Pc – Percentagem de concordância 65%

Tabela 5.11 - Matriz de confusão depois do Agente.

**Operadores/Agente**

	0	1	2	3	4	5	6	7	8	9	10	Oper
0												
1		10	1		1							12
2		4	34	4	1	1						44
3		3	4	49	3		1					60
4			3	4	36	1	2					46
5			1	4	6	42	5	2				60
6					1	1	16	1	1			20
7						1		8				9
8									4			4
9												
10												
AGENTE		17	43	61	48	46	24	11	5			255

Pc – Percentagem de concordância 78%

Foram generalizadas as folhas 50, 453 e 601 com a metodologia aqui apresentada, utilizando este resultado. Este segundo teste, concorda com o primeiro onde a utilização do Agente melhora os resultados dos 3 métodos utilizados. A Rede Neuronal continua a ter um melhor desempenho que as árvores de decisão e de classificação e regressão. Foram efetuados testes qualitativos a todas as folhas e ambos qualitativos e quantitativos à folha 279 cujos resultados se apresentam no ponto seguinte.

### 5.6.3 Validação quantitativa

Para a validação quantitativa, calcularam-se diversas medidas das curvas de nível e compararam-se entre si os valores das curvas homólogas. Utilizaram-se nesta comparação algumas medidas não utilizadas na caracterização numérica das linhas, como a medida da área ocupada pela curva de nível, a diferença das áreas envolvidas e a área diferença, comparou-se ainda a razão entre o comprimento e a raiz quadrada da área. Utilizaram-se neste estudo folhas da série 25k do IGeoE, por conseguinte algumas

curvas de nível são interrompidas nos limites da folha. Nestes casos optou-se por considerar a área formada pela curva de nível e o limite da folha 25k. Os resultados da comparação entre os dados da série 25k e 50k, são apresentados na tabela 5.12, onde o valor das áreas é dado em metros quadrados e o comprimento em metros.

Tabela 5.12 – Comparação entre as curvas de nível originais, escala 25k, e generalizadas manualmente para a escala 50k, folha 279

Curvas de Nível	Area_25k	Comp_25k	$\frac{Comp\ 25k}{\sqrt{Area}}$	Area_50k Manual	Comp_50k	$\frac{Comp\ 50k}{\sqrt{Area}}$
Min	3.0E+03	0.191E+03	2.62	3.1E+03	0.191E+03	2.63
Max	73000.0E+03	130.0E+03	32.08	73000E+03	130.0E+03	31.68
Media	2600.0E+03	7.3E+03	6.02	2650E+03	7.2E+03	5.97
Mediana	84.7E+03	1.6E+03	5.15	86.8E+03	1.6E+03	5.14
DesvPad	8720.0E+03	15.0E+03	3.16	8700E+03	15.0E+03	3.12

**Area\_25k** – Área limitada por uma curva de nível na escala 25k;

**Comp\_25k** – Comprimento total da curva de nível na escala 25k;

$\frac{Comp\ 25k}{\sqrt{Area}}$  – Razão entre o comprimento total da curva de nível na escala 25k e a raiz quadrada da sua área;

**Area\_50k Manual** – Área limitada por uma curva de nível generalizada manualmente, na escala 50k;

**Comp\_50k** – Comprimento total da curva de nível generalizada manualmente, na escala 50k;

$\frac{Comp\ 50k}{\sqrt{Area}}$  – Razão entre o comprimento total e a raiz quadrada da área limitada pela curva de nível, na escala 50k;

Tabela 5.12a – Diferenças entre as curvas de nível originais, escala 25k, e generalizadas manualmente para a escala 50k, folha 279

Curvas de Nível	Dif Area	Dif $\frac{Comp}{\sqrt{Area}}$	Slivers	Slivers-Dif_Area
Min	-23.0E+03	-0.26	0.038E+03	-320.0E+03
Max	340.0E+03	9.16	120.0E+03	150.0E+03
Media	1.8E+03	0.08	6.1E+03	4.3E+03
Mediana	0.058E+03	0.02	1.2E+03	1.3E+03
DesvPad	21.8E+03	0.58	14.0E+03	25.0E+03

**Dif Area** – Diferença entre as áreas limitadas pela curva de nível na escala 25k e da curva respetiva generalizada manualmente na escala 50k;

**Dif  $\frac{Comp}{\sqrt{Area}}$**  – Diferença entre os valores  $\frac{Comp\ 25k}{\sqrt{Area}}$  e  $\frac{Comp\ 50k}{\sqrt{Area}}$ ;

**Slivers** – Área compreendida entre as curvas de nível na escala 25k e a curva de nível respetiva generalizada manualmente na escala 50k;

**Slivers-Dif\_Area** – Diferença entre os valores da área dos Slivers e Dif Area;



Da análise das tabelas 5.12 e 5.12a podemos verificar que o valor da área limitada pela curva de nível diminui genericamente, da 25k para a 50k, aumentando nalguns casos pontuais, principalmente onde o comprimento das linhas é menor e se pretende que as mesmas tenham representatividade nas escalas menores, evitando desta forma que sejam eliminadas. A área dos *slivers*, não correspondendo á diferença das áreas envolvidas, não apresenta um valor muito elevado, pelo que o afastamento das linhas não deve ser muito acentuado. Comparando os valores dados pelo  $Comp/\sqrt{Área}$ , podemos verificar que os valores são semelhantes, sendo inferiores para a série 50k, o que concorda com os dados apresentados anteriormente. Também o comprimento das linhas diminui com a generalização manual, da escala 25k para 50k. O diferencial entre a área dos *slivers* e a diferença entre as áreas ocupadas pela curva nas escalas 25k e 50k não apresentando um padrão, é genericamente positiva, apresentando uma grande dispersão.

Tabela 5.13 – Comparação entre as curvas de nível na escala 50k, generalizadas manualmente e automaticamente, folha 279

Curvas de Nível	Area_50K Manual	Comp_50k Manual	$\frac{Comp\ 50kM}{\sqrt{Área}}$	Area_ Gen Aut	Comp_ GenAut	$\frac{Comp\ 50kGen}{\sqrt{Área}}$
Min	3.1E+03	0.191E+03	2.63	2.9E+03	0.1752E+03	2.62
Max	73000.0E+03	130.0E+03	31.68	73000.0E+03	120.0E+03	29.70
Media	2650.0E+03	7.2E+03	5.97	2500.0E+03	6.74E+03	5.70
Mediana	86.8E+03	1.6E+03	5.14	82.0E+03	1.45E+03	4.99
DesvPad	8700.0E+03	15.0E+03	3.12	8700.0E+03	14.2E+03	2.90

**Area\_50k Manual** – Área limitada por uma curva de nível generalizada manualmente, na escala 50k;

**Comp\_50k Manual** – Comprimento total da curva de nível generalizada manualmente, na escala 50k;

$\frac{Comp\ 50kM}{\sqrt{Área}}$  – Razão entre o comprimento total da curva de nível generalizada manualmente, na escala 50k, e a raiz quadrada da sua área;

**Area\_ Gen Aut** – Área limitada por uma curva de nível generalizada automaticamente, na escala 50k;

**Comp\_ GenAut** – Comprimento total da curva de nível generalizada automaticamente, na escala 50k;

$\frac{Comp\ 50kGen}{\sqrt{Área}}$  – Razão entre o comprimento total e a raiz quadrada da área limitada pela curva de nível generalizada automaticamente, na escala 50k;

Tabela 5.13a – Diferenças entre as curvas de nível na escala 50k, generalizadas manualmente e automaticamente, folha 279

Curvas de Nível	Dif Area	$\text{Dif } \frac{\text{Comp}}{\sqrt{\text{Area}}}$	Slivers	Slivers-Dif Area
Min	-1300.0E+03	-8.45	0.225E+03	0.0024E+03
Max	84.0E+03	1.99	240.0E+03	1300.0E+03
Media	-2.6E+03	0.24	18.0E+03	20.0E+03
Mediana	1.5E+03	0.20	4.9E+03	2.3E+03
DesvPad	86.0E+03	0.64	32.0E+03	94.0E+03

**Dif Area** – Diferença entre as áreas limitadas pela curva de nível generalizada manualmente e automaticamente na escala 50k;

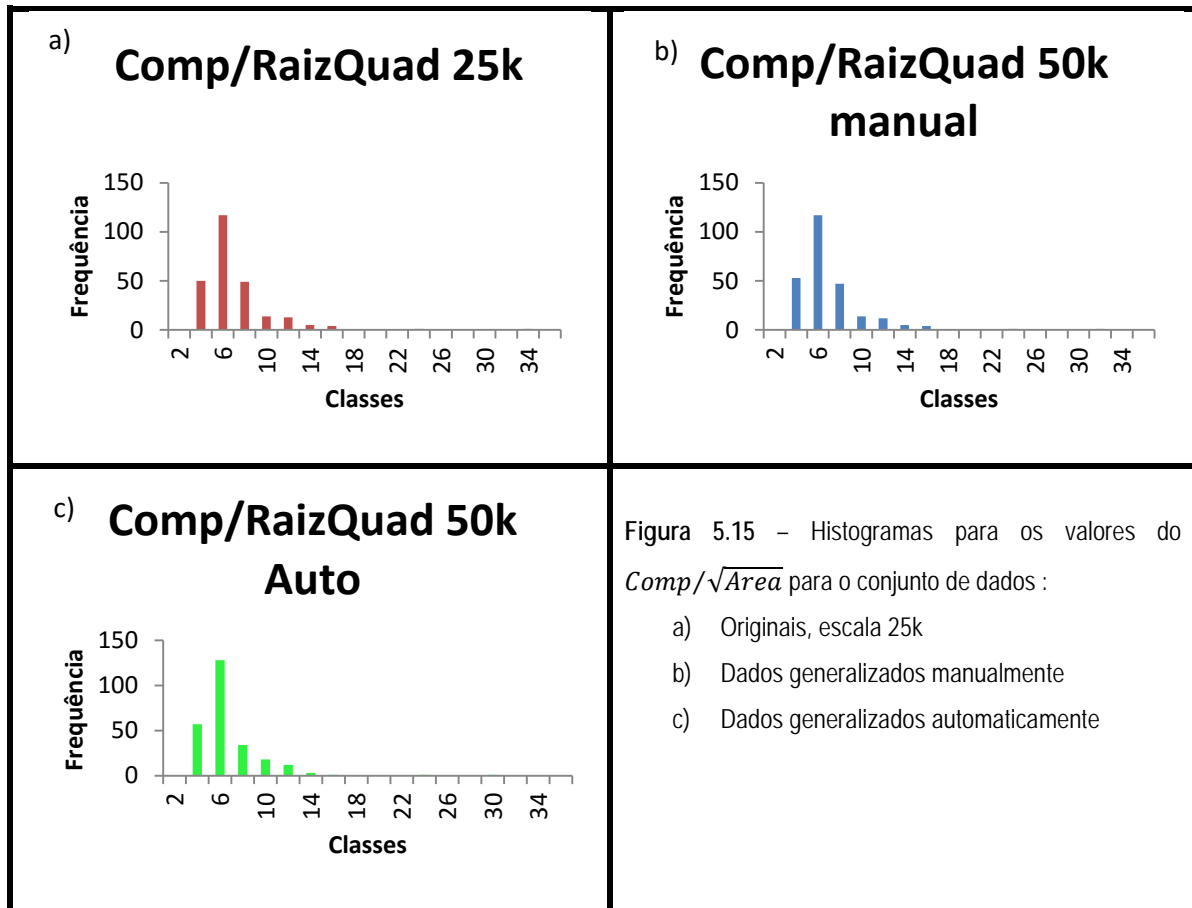
**Dif  $\frac{\text{Comp}}{\sqrt{\text{Area}}}$**  – Diferença entre os valores  $\frac{\text{Comp } 50k}{\sqrt{\text{Area}}}$  e  $\frac{\text{Comp } 50k\text{Gen}}{\sqrt{\text{Area}}}$ ;

**Slivers** – Área compreendida entre as curvas de nível generalizadas manualmente e automaticamente na escala 50k;

**Slivers-Dif\_Area** – Diferença entre os valores da área dos Slivers e Dif Area, na escala 50k;

Comparando as áreas limitadas pelas curvas de nível, generalizadas manualmente e automaticamente (tabela 5.13 e 5.13a), podemos constatar que concordam entre si, nos seus valores máximos e média respetiva, os valores mínimos apresentam uma pequena diferença não muito significativa, devido ao facto do algoritmo não efetuar exageros, diminuindo o valor dos comprimentos e das áreas, mesmo das curvas de nível de menor comprimento. Verifica-se que a diferença entre os  $\text{Comp}/\sqrt{\text{Area}}$  das curvas generalizadas manualmente e automaticamente tem de média um valor próximo de 0.2, os valores máximo, mínimo e mediana também têm valores baixos na ordem 0.2 ou 0.4, havendo concordância entre os valores dados para as curvas de nível generalizadas manualmente e as generalizadas automaticamente com este algoritmo, recorrendo a esta metodologia.

Como podemos verificar nos histogramas da figura 5.15, as linhas têm o mesmo comportamento, tendo os menores valores as linhas correspondentes às curvas de nível generalizadas automaticamente e os maiores as linhas originais na escala 25k. Comparando os histogramas da figura 5.15 a, b e c, podemos verificar que a tendência se mantém havendo coerência nos valores da variável  $\text{Comp}/\sqrt{\text{Area}}$ . Constata-se que cerca de 45% dos elementos estão na 3ª classe, para qualquer dos casos, dados originais ou generalizados. As 2ª, 3ª e 4ª classes englobam cerca de 90% dos casos. Para que o valor do comprimento da linha seja superior dever-se-á generalizar as curvas de nível com valores de tensão inferiores, mas por outro lado o valor da área ocupada pela linha poderá ter um comportamento diferente. Contudo os valores de tensão poderão variar de operador para operador, obtendo-se também soluções corretas.



Também a dimensão fractal sofre uma diminuição com a generalização (tabela 5.14), sendo mais acentuada para o caso da generalização automática, embora não muito significativa, concordando com o resultado apresentado por Lopes (2006). As curvas de nível desta folha apresentam uma dimensão fractal que varia de 1.1 a 1.65 sendo abrangente nas suas formas, a mediana situa-se no 1.37, também a média é 1.36, apresentando este conjunto de valores uma distribuição próxima da normal.

Tabela 5.14 – Comparação entre a dimensão fractal das curvas de nível da folha 25k e 50k, folha 279

	25k	50k	Dif 50k-25k
<b>Min</b>	1.11	1.17	-0.12
<b>Max</b>	1.65	1.56	0.003
<b>Media</b>	1.36	1.34	-0.013
<b>Mediana</b>	1.369	1.337	-0.008
<b>DesvPad</b>	0.08	0.08	0.02

Comparando a área dos *slivers* entre as curvas de nível na escala 25k e 50k generalizadas manual e automaticamente, podemos verificar que a área dos *slivers* é superior no caso das curvas generalizadas automaticamente, o que está de acordo com a

diminuição do comprimento e da dimensão fractal destas curvas de nível. Podemos ainda verificar pelo gráfico da figura 5.16, que se mantém coerentes as áreas dos *slivers* entre a 25k e as 50k generalizadas manual e automaticamente.

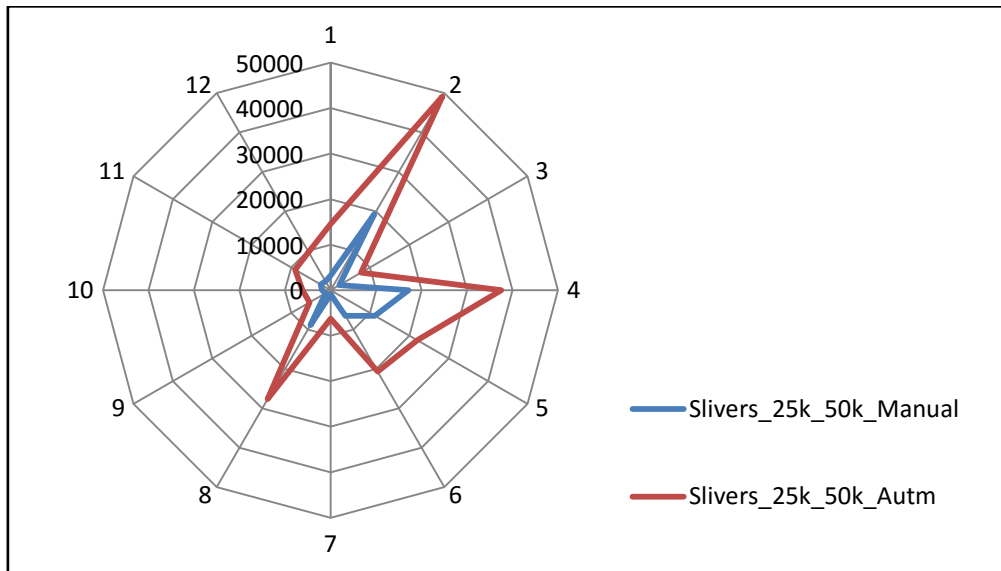


Figura 5.16 – Comparação entre os valores dos slivers, para o conjunto de dados originais e os correspondentes, generalizados manual e automaticamente. As curvas de nível foram agrupadas em 12 conjuntos, representados pelos raios no gráfico.

Comparando o comprimento das curvas de nível generalizadas manual e automaticamente, conforme figura 5.17, podemos verificar que o comprimento das linhas generalizadas manualmente é ligeiramente superior ao comprimento das linhas generalizadas automaticamente. Podemos ainda constatar que não há grandes discrepâncias nas diferenças destes comprimentos, mantendo-se a similaridade das barras do gráfico nas diversas classes, ver figura 5.17.

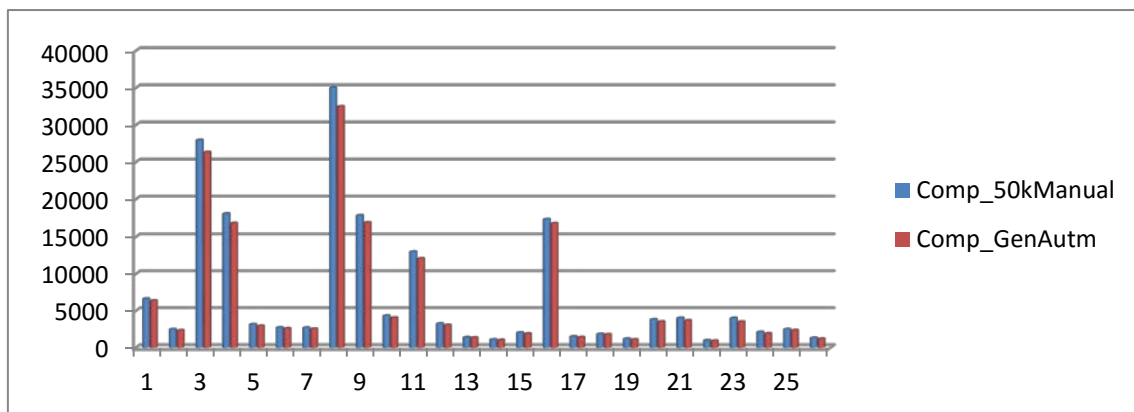


Figura 5.17 – Comparação entre os valores do comprimento, para o conjunto de dados generalizados manualmente e automaticamente.

Tabela 5.15 – Comparação entre os comprimentos das curvas de nível da folha 25k e 50k generalizadas manualmente e automaticamente, folha 279

	25K	50K Gen manual	50k Gen Autom
<b>Min</b>	190.7	190.9	175.1
<b>Max</b>	1.3E+05	1.3E+05	1.2E+05
<b>Media</b>	7.3E+03	7.2E+03	6.7E+03
<b>Mediana</b>	1.58E+03	1.57E+03	1.45E+03
<b>Desv Pad</b>	1.5E+04	1.5E+04	1.4E+04

Verificamos que as diferenças entre o comprimento das curvas de nível generalizadas manualmente e automaticamente é positivo na generalidade dos casos, excetuando-se um número muito limitado de linhas. Desta forma verifica-se que o comprimento das linhas generalizadas automaticamente diminui, relativamente às generalizadas manualmente e a diferença é superior entre as linhas originais e as generalizadas automaticamente, ou seja esta metodologia com este algoritmo deforma mais a curva de nível que a generalização manual, para além disso não efetua exageros.

Após aplicação da metodologia apresentada a 3 folhas da série M888, escala 25k, num total de 606 linhas. Foram as mesmas generalizadas por 3 operadores diferentes, os resultados da comparação entre as tensões obtidas de forma automática e as tensões propostas pelos operadores, são apresentadas na matriz de confusão da tabela 5.16.

Tabela 5.16 – Matriz de confusão das restantes folhas utilizadas

<b>Operadores/Agente</b>											
	0	1	2	3	4	5	6	7	8	9	10
0											
1	3	28	14	2	1	1					
2		2	31	7	4	2					
3			8	65	31	10	6	1	1		
4		1	2	2	67	15	8	1			
5				4	10	170	29	3	1		
6					1	6	34	4	2		
7							1	13	1		
8							1	1	11	1	
9											
10											
AGENTE 3 31 55 80 114 204 79 23 16 1 606											
Pc – Percentagem de concordância 70%											

A percentagem de concordância entre os valores propostos pelos operadores e os calculados pelos métodos de IA e após a utilização do Agente é de cerca de 70%. Os valores distribuem-se em torno da diagonal principal da matriz, significando haver uma

concordância relativa. Como vimos do antecedente os operadores não concordam entre si na escolha do parâmetro a utilizar, o que significa que mesmo utilizando os valores próximos dos valores concordantes, o resultado ainda é aceitável.

## **5.7 Contextualização em Generalização**

### **5.7.1 Generalização contextual entre curvas de nível e pontos de cota**

No processo de generalização cartográfica as entidades geográficas não podem ser consideradas isoladamente sob pena de não serem respeitadas as regras de representação cartográfica. A forma como os objetos são processados depende claramente do seu contexto espacial. A dificuldade associada à contextualização da generalização reside no facto dos algoritmos serem insensíveis à topologia e aplicarem os mesmos critérios independentemente do contexto geográfico. O cartógrafo tem deste modo que verificar e avaliar os resultados intermédios, intervindo no processo sempre que necessário. As situações em que estes conflitos ocorrem são tão variadas, que são de difícil sistematização e assim a sua resolução é ainda hoje estudada. São várias as propostas apresentadas, com as suas vantagens e desvantagens associadas, resultando frequentemente para um número limitado de casos.

A representação do espaço geográfico, a simbolização dos objetos e a escala de representação, originam conflitos gráficos onde é necessário intervir e atuar de forma a obter um mapa legível. Com esta contextualização estamos a generalizar objetos, tendo em conta outros objetos de classes diferentes o que aproxima esta metodologia mais à forma como o operador cartógrafo efetua a generalização.

Com a metodologia anteriormente apresentada pode acontecer que, com a tensão aplicada à linha, a curva generalizada passe para além do ponto de cota, havendo aqui uma nítida violação de topologia. Num primeiro passo, para resolver esta violação é necessário efetuar uma *query* espacial para detecção da violação local. Ou seja, se o ponto de cota estiver no interior da área definida pelas duas curvas de nível, original e generalizada, limitadas pelos pontos médios  $M_i$  e  $M_{i+1}$ , então temos um problema de violação de topologia.

Na correção deste problema, sugere-se que seja aplicada uma tensão inferior nesta parte da curva, entre os dois pontos médios, de forma a curva não passar para além do ponto

de cota. Este é tipicamente um problema a resolver de uma forma recursiva, ou seja, aplicar esta metodologia até que o ponto de cota passe para o exterior da área considerada. Na figura 5.18 d), temos um exemplo desta ação, onde as áreas mais afastadas da curva de nível original representam as sucessivas aplicações de uma tensão inferior. As áreas representadas a vermelho indicam uma violação de topologia. É aplicada sucessivamente uma tensão inferior, até o ponto de cota estar no exterior desta área, representada neste caso, a verde.

Esta metodologia poderá ser igualmente aplicada a violações de topologia entre curvas de nível, evitando a intersecção de diferentes curvas nível entre si ou mesmo a auto intersecção. Ou seja, se no interior da área definida pela linha original e generalizada estiver um ou mais pontos pertencentes a curvas de nível, então aplicar uma tensão inferior. É possível desta forma resolver problemas de topologia, quer com pontos de cota, quer com vértices geodésicos, ou mesmo com as próprias curvas de nível.

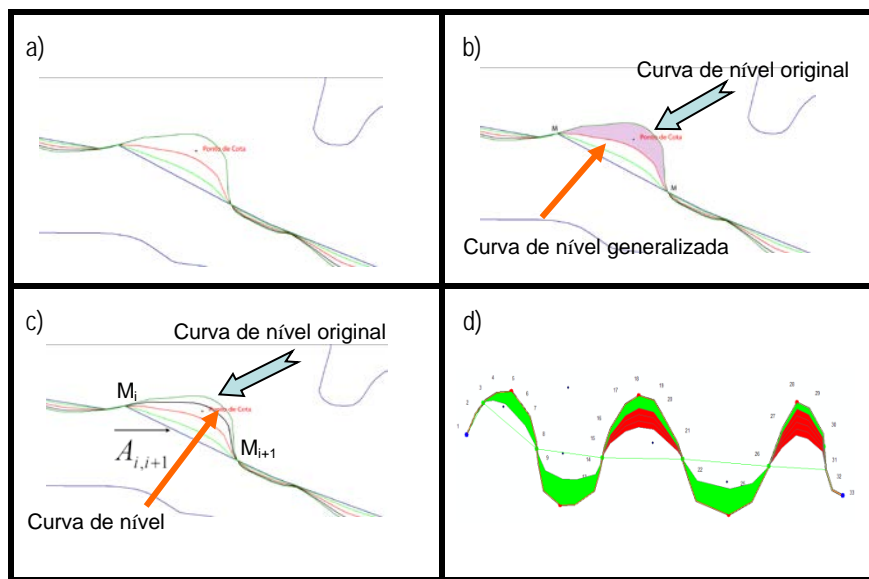


Figura 5.18 – a) Tensão aplicada á linha;  
 b) Query espacial, para detecção da violação da topologia;  
 c) Violação de topologia corrigida, menor tensão aplicada;  
 d) Aplicação de uma tensão inferior, resolução da violação de topologia.

### 5.7.2 Generalização contextual entre curvas de nível e linhas de água

Um segundo problema topológico ocorre quando se efetua a generalização de uma curva de nível, particularmente no troço entre os pontos médios  $M_i$  e  $M_{i+1}$ , onde a tensão a aplicar à linha é o resultado da metodologia apresentada anteriormente. Este é o caso

da curva de nível intersectar uma linha de água, entre os pontos  $M_i$  e  $M_{i+1}$ , então o ponto de intersecção será um máximo local, mais precisamente o vector  $B_j$  entre o vector  $\overrightarrow{A_{i,i+1}} = \overrightarrow{M_i M_{i+1}}$ , conforme figura 5.19, e o ponto de intersecção entre a curva de nível e a linha de água, será o vector  $B_j$  de maior norma entre os pontos  $M_i$  e  $M_{i+1}$ . Segundo as leis de Brisson “uma curva de nível ao atravessar uma linha de água sofre uma inflexão, voltando a convexidade para montante” (Alves et al. 1984). Após a aplicação do algoritmo de generalização, o novo ponto de intersecção entre a linha de água e a curva de nível generalizada, poderá não ser o máximo, se a linha de água não for paralela aos vectores  $B_j$ , perpendiculares ao vector  $\overrightarrow{A_{i,i+1}} = \overrightarrow{M_i M_{i+1}}$ , desta forma torna-se necessário que o ponto de intersecção da linha de água com a curva de nível generalizada, seja um máximo local, de modo a não violar as leis de Brisson.

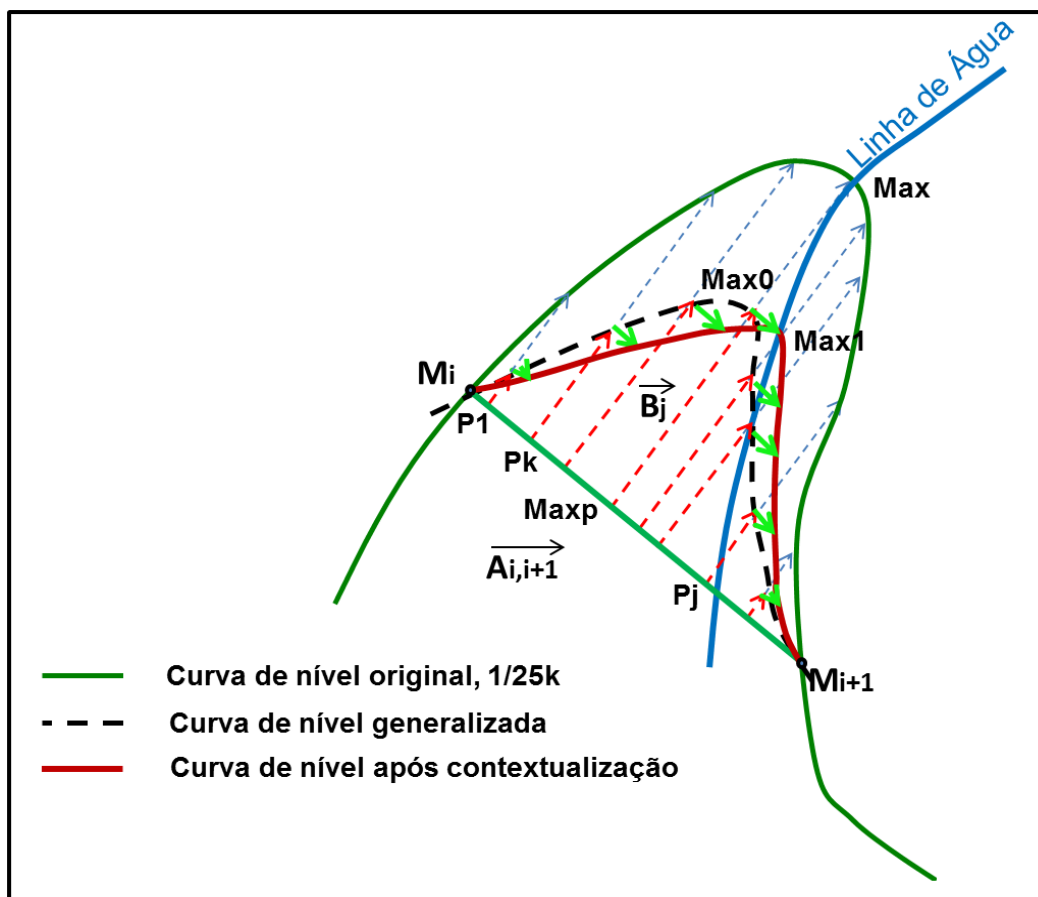


Figura 5.19 – Contextualização da generalização das curvas de nível com as linhas de água

Neste caso o algoritmo de generalização é adaptado para esta situação concreta, executando os passos seguintes, como se pode ver na figura 5.19:



- 1º passo: Executar uma Query Espacial para averiguar a existência de uma intersecção entre o troço da curva de nível e a linha de água, será o ponto Max;
- 2º passo: Aplicar a tensão calculada pela metodologia aqui apresentada; (O ponto Max é projetado em Max0);
- 3º passo: Traçar uma paralela a  $\overrightarrow{A_{i,j+1}}$ , em Max0;
- 4º passo: Executar uma Query Espacial para averiguar a existência de uma intersecção entre a reta paralela a  $\overrightarrow{A_{i,j+1}}$  passando em Max0, e a linha de água, o ponto resultante será o ponto Max1;
- 5º passo: Calcular o deslocamento  $\overrightarrow{Max_0Max_1}$ , paralelo a  $\overrightarrow{A_{i,j+1}}$ ;
- 6º passo: Aplicar aos pontos da curva generalizada os deslocamentos proporcionais a  $\overrightarrow{Max_0Max_1}$ , da seguinte forma, ver figura 5.19:

$$\begin{cases} deslocação\_k = \frac{\|\overrightarrow{M_iP_k}\|}{\|\overrightarrow{M_iMax_p}\|} \times \overrightarrow{Max_0Max_1} \\ deslocação\_j = \frac{\|\overrightarrow{P_jM_{i+1}}\|}{\|\overrightarrow{Max_pM_{i+1}}\|} \times \overrightarrow{Max_0Max_1} \end{cases} \quad (5.2)$$

Desta forma o deslocamento do ponto é máximo no ponto de intersecção da curva de nível com a linha de água, reduzindo o seu valor à medida que nos aproximamos dos pontos médios  $M_i$  e  $M_{i+1}$ , sendo nulo nesses pontos.

Os resultados desta metodologia, generalização das curvas de nível e classificação e seleção das linhas de água para a escala 50k, podem ser vistas na figura 5.20.

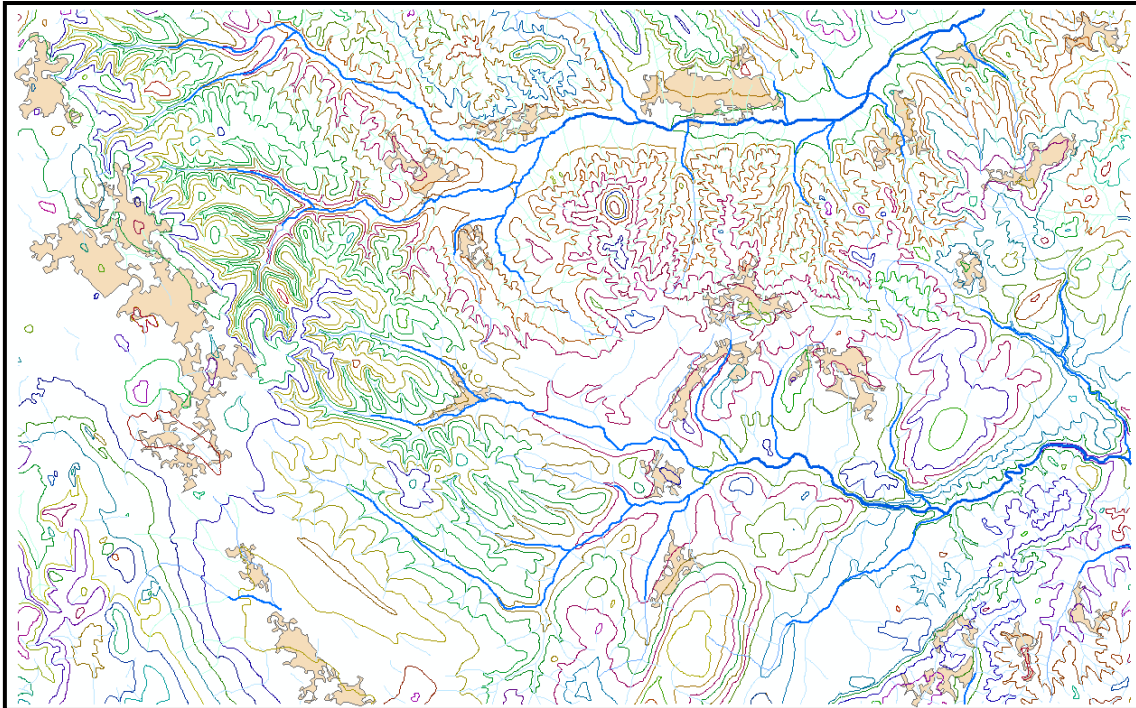


Figura 5.20 – Resultados da metodologia aplicada a curvas de nível e linhas de água

## 5.8 Comparação com outros métodos

A generalização contextual para classes de objetos individuais foi relatada por diversos autores e.g. (Bader et al. 2005; Mustière e Moulin, 2002; Basaraner e Selcuk 2008, Steiniger et al. 2010; Ilgin et al. 2006; Daoud e Doytscher, 2008). Contudo, nenhum destes métodos considera a relação semântica ou espacial entre classes de objetos diferentes num mapa, nem um método para formalizar ou representar tais relações. Uma tentativa de tratar simultaneamente classes de objetos diferentes foi apresentada por Gaffuri (2007, 2008), no modelo GAEL (*Generalisation based on Agents and Elasticity*), ver sobre este assunto (Steiniger e Weibel 2007).

Nesta tese propõe-se uma generalização contextual, utilizando os dados a 2D, do relevo, ou seja as curvas de nível, e a hidrografia onde as suas relações espaciais e topológicas são preservadas após a generalização. O modelo GAEL necessita de construir um modelo digital do terreno no formato TIN e é a partir desse modelo que calcula o parâmetro indicador de drenagem que se baseia no ângulo entre dois vectores: o vector tangente à linha no ponto considerado e a inclinação do vector da superfície. O modelo GAEL permite também a visualização das partes da rede hidrográfica que têm problemas de drenagem. A metodologia utilizada no modelo GAEL, pode ser usada independentemente do processo de generalização, assim é possível detetar os

confinamentos eventuais que possam existir entre a rede hidrográfica e o relevo. A classe dos confinamentos força cada triângulo do agente relevo, movendo-o e rodando-o, para melhorar o estado de satisfação das linhas hidrográficas, que estão sobre o relevo. O objetivo é tentar encontrar um equilíbrio entre o relevo e as linhas de água, com a possibilidade de remodelação dos objetos do relevo ou de ambos.

O modelo GAEL impõe assim a colocação das linhas de água nos seus talvegues, movendo a linha de água ou recalculando o DTM. A última etapa do modelo GAEL é criar um novo DTM (modelo TIN) com as curvas de nível generalizadas, a fim de criar triângulos novos, de modo que as linhas de água possam contextualizar com as curvas de nível.

A abordagem aqui apresentada é conceptualmente similar ao modelo GAEL usando as linhas de água como um confinamento à curva de nível generalizada, ou como um confinamento ao modelo digital do terreno, como é o caso do projeto (GAEL). Embora haja diferenças significativas na execução: a) o projeto GAEL utiliza os dados 3D, esta abordagem usa dados 2D; b) o modelo GAEL usa os pontos das curvas de nível como agentes para calcular a sua nova posição, e o método aqui proposto usa um agente para seleccionar o melhor parâmetro, entre os três métodos de IA, para generalizar a linha com o algoritmo. c) O modelo GAEL recalcula o DTM para gerar de novo as curvas de nível, com o custo desta transformação associado. A abordagem aqui apresentada não utiliza o DTM, evitando assim cálculos intermédios.

A utilização de modelos 3D foi experimentada por Zhilin Li (Li et al. 1999), a sua metodologia englobava a construção de um DTM a partir das curvas de nível, seguindo-se a generalização desse DTM e posteriormente a extração de curvas de nível a partir do DTM generalizado. É suposto que no final deste processo as curvas de nível estejam generalizadas para a escala pretendida. A vantagem desta abordagem é que haverá uma garantia de não haver interseções entre as curvas de nível se forem utilizados algoritmos apropriados de interpolação de curvas de nível. Após a generalização do DTM, poderá haver inconsistências nas curvas de nível resultantes da sua vectorização, resultando em erros topológicos, causados pelo *resampling* ou reamostragem. Na maioria dos casos é aplicado um filtro passa-baixo para suavizar o DTM. Uma outra desvantagem desta abordagem é que uma perda na exatidão e/ou na fidelidade pode ser causada e pode ser introduzido ruído durante o processo de conversão de curva de nível/DTM/curva de

nível. A generalização de curvas de nível em 2D aproxima-se mais da forma como o operador cartógrafo executa esta tarefa. Contudo nenhum dos métodos dispensa a avaliação final efetuada pelo ser humano.

# Capítulo 6

## Conclusão

---

Na atualidade, a tendência predominante na indústria cartográfica consiste no múltiplo uso dos dados armazenados em bases de dados geoespaciais. A obtenção e manutenção de dados geoespaciais são atividades extremamente dispendiosas para as agências cartográficas nacionais.

O processo de generalização cartográfica como processo integrante de cadeias de produção das agências cartográficas nacionais decorre dos compromissos legais e institucionais de produção e preservação de cartografia multi-escala. Ao longo do tempo foram propostas diversas abordagens para resolução deste problema com soluções de complexidade crescente, desde soluções pontuais com operadores de simplificação e suavização de linhas, a sistemas inteligentes com recurso a agentes e constrangimentos multicritério. Embora a generalização em ambiente digital tenha muitas diferenças do processo manual de generalização tradicional, as razões principais para generalizar continuam a ser as mesmas. A generalização é motivada pela necessidade de fornecer múltiplas vistas de dados geográficos em várias escalas e níveis de definição.

O resultado da generalização cartográfica poderá ser utilizado não só para a produção de mapas mas também para visualização de dados em monitores de navegação, integrar BDG de várias resoluções ou mesmo permitir legibilidade em dados geográficos, transferência de dados via internet ou outras aplicações como consultas de bases de dados geográficos ou produção de mapas *on-demand* a utilizar em zonas de conflitos militares, ou missões de apoio às populações em caso de desastres naturais ou tecnológicos de grande dimensão.

Nesta dissertação apresentámos e debatemos os tópicos atuais da generalização como a automatização do processo de generalização como um processo cognitivo em que a máquina procura substituir o processo humano de visão e raciocínio de forma inteligente com recurso a algoritmos de inteligência artificial. Neste contexto, foi proposto um novo algoritmo de generalização cartográfica de linhas, aplicado a curvas de nível. A parametrização do processo de generalização da linha é efetuado de forma automática com base em algoritmos de inteligência artificial, um agente e uma base de dados do conhecimento. São também apresentadas abordagens ao problema da contextualização da generalização cartográfica de curvas de nível com pontos cotados e linhas de água. São apresentados exemplos de aplicação do algoritmo no processo de generalização da cartografia 25k para 50k, a sua avaliação quantitativa e ainda avaliação qualitativa efetuada por operadores cartógrafos.

Esta metodologia apresenta como vantagens, a generalização de curvas de nível em 2D. A abordagem apresentada está mais próxima da execução manual efetuada pelo cartógrafo e do seu conhecimento em generalização. A utilização do algoritmo é intuitiva, utilizando apenas um parâmetro, permitindo ao cartógrafo antever resultados. Os algoritmos de generalização de linhas, mostram-se pouco eficientes neste caso, pois não têm em consideração a contextualização com outros temas. Esta abordagem contrasta com algumas propostas, que utilizam o MDT. Outra das vantagens é a escolha do parâmetro a utilizar no algoritmo, efetuada pelo computador e dependendo da caracterização intrínseca de cada curva de nível, adaptando a generalização á curva particular. Após a generalização, a curva de nível não necessita de uma suavização adicional.

É possível utilizar a IA para algumas tarefas em generalização, podendo neste caso substituir o cartógrafo na decisão da escolha da tensão a utilizar na generalização de uma curva de nível, após a sua caracterização numérica. Contudo o valor de tensão poderá não ser apropriado para a totalidade da curva de nível, nesse caso poder-se-ia utilizar um método de segmentação, já apresentado por outros autores e, aplicar esta metodologia aos segmentos respetivos.

Há um ganho de produtividade avaliado em 3 a 4 dias de trabalho por carta 50k. Após o treino dos métodos de IA, o tempo gasto pelo operador cartógrafo é reduzido nesta tarefa em cerca de 80%. De salientar que o treino dos métodos de IA, só é efetuado na

implementação da metodologia, após estar em produção poderá haver necessidade de treinar de novo os métodos de IA caso haja casos de curvas de nível que ainda não foram tratados.

Depois da seleção dos pontos de cota, vértices geodésicos e classificação e seleção das linhas de água a representar na 50k, foi apresentada uma metodologia para efetuar a contextualização das curvas de nível com estes temas, evitando violações topológicas. Apesar de haver uma avaliação e correções finais a realizar pelo operador, com estes automatismos ganha-se tempo significativo de operador e consistência no processo.

Da análise comparativa e sensitiva com cartas de referência verificou-se que:

- O atributo mais significativo para a caracterização numérica da forma das curvas de nível, foi a sua dimensão fractal.
- A dimensão fractal diminui com a generalização, sendo esta variação superior para a generalização automática.
- Os operadores cartógrafos não concordam entre si quanto ao valor da tensão a utilizar, para cada linha.
- O nível de concordância entre a tensão utilizada pelo operador cartógrafo e os métodos de IA, aumenta significativamente quando se utiliza o Agente.
- A área limitada pelas curvas de nível diminui genericamente com a generalização.
- O comprimento das linhas diminui com a generalização, sendo esta diminuição superior no caso da generalização automática.
- O diferencial entre a área dos *slivers* e a diferença das áreas (25k-50k), não apresenta um padrão, no entanto é genericamente positivo.
- A área dos *slivers* (25k-50k), é superior para o caso das linhas generalizadas automaticamente, mantendo a coerência com a área dos *slivers* (25k-50k) com as linhas generalizadas manualmente.

Uma das desvantagens do algoritmo é que o mesmo não se adapta á generalização de edifícios ou aglomerados urbanos.

As contribuições dadas por esta dissertação são passíveis de utilizar noutras classes de objetos, ou temas.





## Referências Bibliográficas

- Ai, T., (2004), “A displacement of building cluster based on field analysis”, *Acta Geodaetica et Cartographica Sinica*, 33, 89-94.
- Alves, D., Cruz, J., e Norte, C., (1984), “Topografia”, Manual da cadeira 23E, Academia Militar, Lisboa.
- Aoyama, H., e Kawagoe, M., (1991), “A piecewise linear approximation method preserving visual feature points of original figures”, *CVGIP: Graphic Models and Image Processing*, 3(5), 435-146.
- Bader, M., e Weibel, R., (1997), “Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal Maps”, In: *Proceedings of the 18<sup>th</sup> ICA/ACI International Cartographic Conference*, Stockholm, 1997. 1525-1534p.
- Bader, M., e Barrault, M., (2000), “Improving Snakes for Linear Feature Displacement in Cartographic Generalization”, *Spatial Data Handling*, Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland;
- Bader, M., e Barrault, M., (2001), “Cartographic Displacement in Generalization: Introducing Elastic Beams”, *Fourth Workshop on Progress in Automated Map Generalization*, Beijing, China, August 2-4.
- Bader, M., Barrault, M., Regnault, N., Mustière, S., Duchêne, C., Ruas, A., Fritsch, E., Lecordix, F., e Barrillot, X., (1999), “AGENT Workpackage D2 - Selection of Basic Algorithms”. Technical report, University of Zurich.
- Bader, M., Barrault, M., e Weibel, R., (2005), “Building Displacement over a Ductile Truss”, *International Journal of Geographical Information Science*, 19(8-9), 915-936.
- Balboa, J., e López, F., (2000), “Frequency filtering of linear elements by means of wavelets - A method and example”, *The Cartographic Journal*, 37(1), 39-50.
- Balboa, J., e López, F., (2008), “Generalization-oriented Road line classification by Means of an Artificial Neural Network”, in *Vários, Geoinformatica*, Grupo de Investigación en Ingeniería Cartográfica, Dpto. de Ingeniería Cartográfica,

- Geodésica y Fotogrametría, Universidad de Jaén. Campus, Las Lagunillas, s/n. 23071, Jaén (Spain), p 289-312.
- Balboa, J., e López, F., (2009), “Sinuosity pattern recognition of road features for segmentation purposes in cartographic generalization”, Grupo de Investigación en Ingeniería Cartográfica, Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén. Campus, Las Lagunillas, s/n. 23071, Jaén (Spain).
- Balboa, J., López, F., e Luque, R., (2008), “Road line classification for cartographic generalization: a neural net approach”, Grupo de Investigación en Ingeniería Cartográfica, Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén. Campus, Las Lagunillas, s/n. 23071, Jaén (Spain).
- Barrault, M., Regnaud, N., Duchêne, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A., e Weibel, R., (2001), “Integrating multi-agent, object-oriented, and algorithmic techniques for improved automated map generalization”, proceedings of the 20th international conference of cartography, ICA, Beijing, China, volume 3, pp 2110-2116.
- Basaraner, M., e Selcuk, M., (2008), “A Structure Recognition Technique in Contextual Generalisation of Buildings and Built-up Areas”, The Cartographic journal Vol. 45 No. 4 pp. 274–285 November 2008, © The British Cartographic Society 2008.
- Battersby, S., e Clarke, K., (2003), “Quantifying Information in Line Generalization”, Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106 USA, Proceedings of the 21<sup>st</sup> International Cartographic Conference (ICC), Durban, South Africa, 10-16 August. pp. 118-126.
- Beale, R., et al., (1990), “Neural Computing: An Introduction.”, Adam Hilger, UK, 1990.
- Beard, M., (1991), “Theory of a Cartographic Line Revisited: Implications for Automated Generalization”, Cartographica, Vol. 28, No. 4, pp. 32-58.
- Berg, M., Kreveld, M., e Schirra, S., (1995), “A new approach to subdivision simplification”, In Twelfth International Symposium on Computer-Assisted Cartography, volume 4, pages 79–88, Charlotte, North Carolina.

- Berg, M., Kreveld, M., e Schirra, S., (1998), “Topologically Correct Subdivision Simplification Using the Bandwidth Criterion”, *Cartography and Geographic Information Systems*, 25(4), 243–257.
- Bench-Capon, T., (1990) “Knowledge representation: an approach to artificial intelligence”, Academic Press Professional, Inc. San Diego, CA, USA ©1990 ISBN:0-12-086440-1.
- Bernhardt, M., (1992), “Quantitative Characterization of Cartographic Lines for Generalization”, Department of Geodetic Science and Surveying, The Ohio State University Columbus, Ohio 43210-1247.
- Bertin, J., (1983), “Semiology of graphics”, University of Wisconsin Press ©1983, ISBN:0299090604, 456p.
- Bertin, J., (1999), “Sémiologie graphique. Les diagrammes, les réseaux, les cartes”, Paris, École Des Hautes Études En Sciences Sociales, 444p.
- Boutoura, C., (1989), “Line generalization using spectral techniques. ”, *Cartographica*, 26(3&4), 33-18.
- Braga, A., Carvalho, A., e Ludemir, T., (2007), “Redes neurais artificiais: teoria e aplicações”, Rio de Janeiro: LTC, 2007. 226p.
- Brassel, K., e Weibel, R., (1988), “A Review and Framework of Automated Map Generalization”, *Int. Journal of Geographical Information Systems*, 2(3): 229-244.
- Brenner, C., e Sester, M., (2005), “Cartographic Generalization using primitives and constraints”, XXII International Cartographic Conference, ICC 2005, ISBN:0-958-46093-0, A Coruña, Spain, 11-16 July 2005, 10p.
- Burghardt, D., (2005), “Controlled line smoothing by snakes”, *GeoInformatica* 9:3, 237-252, 2005.
- Burghardt, D., e Meier, S., (1997), “Cartographic displacement using the snakes concept”, In: Foerstner W. and Pluemer, L. (eds): *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*. Birkhauser-Verlag, Basel, Switzerland, pp. 59-71.

- Buttenfield, B., (1989), "Scale-Dependence and Self-Similarity in Cartographic Lines", *Cartographica*, 26(1): 79-100.
- Buttenfield, B., (1992), "A rule for describing line feature geometry", In B. Buttenfield and R. McMaster, Eds. *Map Generalization*. Hallow, Essex, U.K.: Longman Scientific.
- Buttenfield B., e McMaster R.,(1991), "A rule for describing line feature geometry", In B. Buttenfield and R. McMaster, Eds. *Map Generalization. "Making Rules for Knowledge Representation"* Hallow, Essex, U.K.: Longman Scientific, London.
- Byong-Nam, C., Young-Gul, K., (2007), "Framework and Workflows for Spatial Database Generalization", The Authors. *Journal compilation © 2007 Blackwell Publishing Ltd, Transactions in GIS*, 2007, 11(1): 101–114.
- Cartwright, W., (2011), "Representations, Diagrams and Visualizations of Space and Place", *Lecture notes in Geoinformation and Cartography, Advances in Cartography and GIScience Volume1* 73-90. Springer, 2011.
- Cecconi, A., (2003), "Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping", PhD Thesis, University of Zurich, 138p.
- Cetinkaya, B., Aslan, S., Sengun, Y., e Cobankaya, O., (2006), "Contour Simplification with Defined Spatial Accuracy", Dursun Er Ilgin, General Command of Mapping, 06100 Cebeci, Ankara, Turkey.
- Christ, F., (1978), "A Program for the Fully Automated Displacement of Point and Line Features in Cartographic Generalizations", *Informations Relative to Cartography and Geodesy, Translations*, 35:5- 30.
- Christensen, H., (1999), "Cartographic line generalization with waterlines and medial-axes", *Cartography and Geographic Information Science*, Vol. 26(1):19–32, 1999.
- Christensen, H., (2000), "Line generalization by water lining and a medial-axis transformation. Successes and issues in an implementation of Perkal's proposal", *The Cartographic Journal*, Vol. 37(1):19–28, 2000.

- Costa, E., e Simões, A., (2004), “Inteligência Artificial – Fundamentos e Aplicações”, FCA - Editora de Informática, Lda, ISBN 972-722-269-2, 581p.
- Cromley, R., e Campbell, G., (1992), “Integrating Quantitative and Qualitative Aspects of Digital Line Simplification”, *The Cartographic Journal* 29(1): 25-30.
- D’Algé, J., (2007), “Generalização cartográfica em sistemas de informação geográfica: aplicação aos mapas de vegetação da amazônia brasileira”, Universidade de São Paulo 2007, tese Doutorado.
- Daoud, J., e Doytsher, Y., (2008), “An automated cartographic generalization process: a pseudo-physical model”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol XXXVII. Part B2. Beijing 2008.
- Douglas, D., e Peucker, T., (1973), “Algorithms for the Reduction of the Number of Points Required to Represent a Line or Its Character”, *The American Cartographer* 10(2):112-123.
- Dutton, G., (1999), “A hierarchical coordinate system for geoprocessing and cartography”, *Lecture Notes in Earth Science* 79. Berlin: Springer-Verlag. XIX, 231 pp. ISBN 3-540-64980-8.
- Falconer, K., (1990), “Fractal Geometry, Mathematical foundations and applications”, John Wiley & sons Ltd., Baffins Lane, Chichester, West Sussex PO19 1UD, England.
- Feng, Q., e A-Xing, Z., (2003), “Knowledge discovery from soil maps using inductive learning”, *International Journal of Geographical Information Science*, VOL. 17, NO. 8, 771–795.
- Foerster, T., (2010), “Web-based Architecture for On-demand Maps -Integrating Meaningful Generalization Processing”, PhD thesis, University of Twente, Faculty of ITC, 173p.
- Foerster, T., Stoter, J., e Köbben, B., (2007), “Towards a Formal Classification of Generalization Operators”, In: *Proceedings of the 23th International Cartographic Conference*. Moscow, Russia: 3-10 August.

- Foerster, T., Stoter, J., e Kraak, M., (2010), “Challenges for Automated Generalisation at European Mapping Agencies - A Qualitative and Quantitative Analysis”, *The Cartographic Journal* Vol. 47 No. 1 pp. 41–54 February 2010.
- Gaffuri, J., (2007), “Field Deformation in an Agent-Based Generalisation Model: the GAEL Model”, COGIT laboratory, Institut Géographique National, France.
- Gaffuri, J., (2008), “Généralisation automatique pour la prise en compte de thèmes champ le modèle GAEL”, Université Paris-Est École doctorale ICMS, Institut Géographique National Laboratoire COGIT, PhD These, Paris.
- Galanda, M., (2003), “Automated Polygon Generalization in a Multi Agent System”, Ph.D. thesis, Department of Geography, University of Zurich, 176p.
- Galanda, M., e Weibel, R., (2002), “An Agent-based Framework for Polygonal Subdivision Generalization”, In: Richardson, D, and Van Oosterom, P. (eds., 2002), *Advances in Spatial Data Handling*, Berlin: Springer-Verlag, 121-136.
- Gama et al., (2012), “Extração de Conhecimento de Dados”, *Data Mining*, Edições Sílabo, Lda. Lisboa ISBN 978-972-618-698-4.
- Gaspar, J., (2008), “Dicionário de Ciências Cartográficas”, 2ª edição, Lidel – Edições técnicas, Lda, ISBN 978-972-757-546-6, Lisboa.
- Grünenreich, D., (1985), “Computer-assisted generalization”, In: *Papers CERCO-Cartography Course*. Frankfurt, Germany: Institut fur Angewandte Geodasie.
- Grünenreich, D., (1992), “ATKIS - A topographic information system as a basis for GIS and digital cartography in Germany”, In: R. Vinken (Ed.) *From Digital Map Series to Geo-Information Systems*, Geologisches Jahrbuch Series A. Hannover, Germany: Federal Institute of Geosciences and Resources.
- Grünenreich, D., (1995a), “Development of computer-assisted generalization”, In: J.- C. Muller, J.-P. Langrange, and R. Weibel (Eds.) *GIS and Generalization: Methodology and Practice*. London, England: Taylor & Francis.
- Grünenreich, D., (1995b), “Development of Computer-Assisted Generalization on the Basis of Cartographic Model Theory”, in: *GIS and Generalization – Methodology and Practice*, Great Britain, Taylor&Francis, London, pp. 47-55.

- Harrie, L., (1999), "The Constraint Method for Solving Spatial Conflicts in Cartographic Generalization", *Cartography and Geographic Information Science*, Vol. 26, No. 1, pp. 55-69.
- Harrie, L., e Sarjakoski, T., (2000), "Generalisation of vector data sets by simultaneous least squares adjustment", Department of Cartography and Geoinformatics, Finnish Geodetic Institute, Finland, International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B4. Amsterdam 2000.
- Hetch-Nielsen, R., (1990), "Neurocomputing", Addison-Wesley, Reading, MA, 1990, 433p.
- Højholt, P., (2000), "Solving Space Conflicts in Map Generalisation: Using a Finite Element Method", *Cartography and Geographic Information Science*, Vol. 27, No. 1, pp. 65-74.
- Holsheimer, M., e Siebes, A., (1994), "Data mining: the search for knowledge in databases", Technical Report CS-R9406, CWI, Amsterdam, The Netherlands.
- Hubert, F., e Ruas, A., (2003), "A method based on samples to capture user needs for generalisation", In 5<sup>th</sup> ICA workshop on progress in automated map generalization, Paris, 2003.
- Ilgin D., et al., (2006), "Contextual Generalisation Implementation in Topographic Mapping: KartoGEN Project", ICA Commision on Generalisation and Multiple Representation – Research Workshop, Leicester, UK.
- Jasinski, M., (1990), "The comparison of complexity measures for cartographic lines", Technical Report 90-1, Buffalo, New York: National Center for Geographic Information and Analysis, 65p.
- João, E., (1998), "Causes and Consequences of map Generalisation", Taylor & Francis London School of Economics, 288p.
- Jones, C., Bundy, G., e Ware, J., (1995), "Map Generalization with a Triangulated Data Structure", *Cartography and Geographic Information Systems*, 1995. 22 (4). 317-331p.
- Kazemi, S., Lim, S., e Ge, L., (2005), "Integration of Cartographic Knowledge with Generalization Algorithms", IGARSS 05, Geoscience and Remote Sensing Symposium, 2005. IEEE International, Volume: 5.

- Kilpeläinen, T., (1992a), "Generalization. Not in the Domain of Maps but in the Domain of Geographical Databases", *Surveying Science in Finland*, Vol.10, no. 2, pp. 11-33.
- Kilpeläinen, T., (1992b), "Tree matching problems with applications to structured text databases", Ph.D. Dissertation. Department of Computer Science, University of Helsinki, Helsinki, Finland, 109p.
- Kilpeläinen, T., (1997), "Multiple representation and generalisation of geo-databases for topographic maps", *Publications of the Finnish Geodetic Institute*, Vol. N° 124.
- Kilpeläinen, T., (2001), "Maintenance of multiple representation databases for topographic data", *The Cartographic Journal* 37 (2), 101-107.
- Kohavi, R., e Quinlan, J., (1999), "Decision-tree discovery", In Will Klossgen and Jan M. Zytkow, editors, *Handbook of Data Mining and Knowledge Discovery*, chapter 16.1.3, pages 267-276. Oxford University Press, 2002.
- Ladeira, M., (1997), "Representação de Conhecimento e Redes de Decisão", Tese Doutorado em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Instituto de Informática, Porto Alegre, Março de 1997.
- Lamy, S., et al., (1999), "The Application of Agents in Automated Map Generalisation", Presented at the 19th ICA Meeting Ottawa, Aug 14-21 1999, pp1225-1234.
- Lang, T., (1969), "Rules for robot draughtsmen", *Geographical Magazine*, Vol 42, pp. 50-51.
- Larose, D., (2005), "Discovering knowledge in data, an Introduction to Data Mining", Published by John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN 0-471-66657-2, 336p.
- Lecordix, F., Rousseau, D., e Rousseau, T., (1995), "An evaluation of Map Generalizer, an Intergraph Interactive Generalization Software Package", IGN & OEEPE Report, 30p.
- Lecordix, F., Plazanet, C., e Lagrange, J., (1997), "PlaGe : a Platform for research in Generalization, Application to caricature", Institut Géographique National (France), Research Service, COGIT Laboratory, 2 Avenue Pasteur 94160 Saint-Mandé.



- Leitner, M., e Buttenfield, B., (1995), "Multi-scale Knowledge Acquisition: Inventory of European Topographic Maps", *Cartography and GIS*, 22(3): 232-241.
- Li, Z., (1993), "Some considerations on the issue of line generalization", *The Cartographic Journal*, 30(1): 68-71.
- Li, Z., (2006), "Algorithmic Foundation of Multi-scale Spatial Representation", CRC Press (Taylor & Francis Group), Bacon Raton. pp 280.
- Li, Z., (2007), "Digital Map Generalization at the Age of Enlightenment: a Review of the First Forty Years", *The Cartographic journal* Vol. 44 No. 1 pp. 80-93 February 2007.
- Li, Z., e Openshaw, S., (1992), "Algorithms for objective generalization of line features based on the natural principle", *International Journal of Geographical Information Systems*, 6(5): 373-389.
- Li, Z., Sui, H., e Gong, J., (1999), "A System for Automated Generalisation of Contour Lines", Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong.
- Li, Z., Khoshelham, K., Ding, X., e Zheng, D., (2004a), "Empirical mode decomposition (EMD) transform for spatial analysis", In *Advances in Spatial Analysis and Decision Making*. Z. Li, Q. Zhou and W. Kainz (Eds), pp. 19-29.
- Li, Z., Yan, H., Ai, T., e Chen, J., (2004b), "Automated Building Generalization Based on Urban Morphology and Gestalt Theory", *International Journal of Geographical Information Science*, 18(5): 513-534.
- Lichtner, W., (1979), "Computer-Assisted Processes of Cartographic Generalization in Topographic Maps", *Geo-Processing*, 1:183-199.
- Lonergan, M., e Jones, C., (2001), "An Iterative Displacement Method for Conflict Resolution in Map Generalization", *Algorithmica*, Volume 30, Issue 2, pp 287-301.
- Lopes, J., (2006), "Generalização Cartográfica", Tese de mestrado, Mestrado em Ciências e Engenharia da Terra, Faculdade de Ciências da Universidade de Lisboa, 94p.

- López, J., e Balboa, J., (2007), “Generalization-oriented Road Line Classification by Means of an Artificial Neural Network”, Grupo de Investigación en Ingeniería Cartográfica, Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén. Campus, Las Lagunillas, s/n. 23071, Jaén (Spain), 12: pp 289-312.
- Mackaness, W., (1994), “An Algorithm for Conflict Identification and Feature Displacement in Automated Map Generalization”, *Cartography and Geographic Information Systems*, 21 (4), 1994. 219-232p.
- Mackaness, W., (1995), “Analysis of Urban Road Networks to Support Cartographic Generalization: *Cartography and Geographic Information Systems*”, Vol. 22, pp 306-316.
- Mackaness, W., Ruas, A., e Sarjakoski, L., (2007), “Generalization of geographic information: Cartographic modeling and applications”, International Cartographic Association, ISBN 978-0-08-045374-3.
- Mackaness, W., e Steven, M., (2006), “An Algorithm for Localised Contour Removal over Steep Terrain”, *The Cartographic Journal* Vol. 43 No. 2 pp. 144–156 July 2006.
- Marshall, G., (1990), “Advanced Students' Guide to Expert Systems”, Heinemann Newnes, Oxford, UK, pp. 176, 1990.
- Másson, E., e Wang, Y., (1990), “Introduction to computation and learning in artificial neural networks”, *European Journal of Operational Research*. v. 47, p. 1-28, 1990.
- Mateus, M., (1999), “Carta Militar de Portugal, Série M782: uma carta estratégica”, *Boletim do IGeoE* N°61, Novembro 1999, ISSN 0872-7600.
- Maudet, A., Touya, G., Duchêne, C., e Picault, S., (2014), “Advances in Practical Applications of Heterogeneous Multi-Agent Systems”, *The PAAMS Collection Lecture Notes in Computer Science*, Volume 8473, 2014, pp 355-358.
- McMaster, R., (1986), “A statistical analysis of mathematical measures for linear simplification”, *The American Cartographer*, 13(2): 103-117.
- McMaster, R., (1991), “Conceptual framework for geographical knowledge”, In B.P. Buttenfield, R. McMaster and H. Freeman (Eds.), *Map Generalization: Making*

- Rules for Knowledge Representation, Harlow, Essex, England Longman Scientific & Technical, New York NY: Wiley, 1991, pp. 21-39.
- McMaster, R., (1995), "Knowledge acquisition for cartographic generalization: experimental methods", in J.C. Müller, J.P. Lagrange, and R. Weibel (Eds.), *GIS and Generalization: Methodology and Practice*, London: Taylor & Francis, 1995.
- McMaster, R., e Shea, S., (1992), "Generalization in Digital Cartography". Association of American Geographers, Washington, 134pp.
- Meng, L., (2003), "Cognitive Modelling of Cartographic Generalization", *Strategies on Automated Generalization of Cartographic Data*, Project Report. Retrieved in July 2003, 76p.
- Mitchell, T., (1997), "Machine Learning", McGraw-Hill Science/Engineering/Math; ISBN: 0070428077.
- Müller, J., (1986), "Fractal dimension and inconsistencies in cartographic line representation", *The Cartographic Journal*, 23, 123-130.
- Müller, J., (1990), "The removal of spatial conflicts in line generalization", *Cartography and Geographic Information Science*, 1 April 1990, vol. 17, no. 2, pp. 141-149(9)
- Müller, J., Lagrange, J., Weibel, R., (1995b), "GIS and Generalization – Methodology and Practice", *GISDATA 1*, Taylor & Francis, ISBN 0-7484-0319-1.
- Müller, J., e Mouwes, P., (1990), "Knowledge acquisition and representation for rule based map generalization: An example from the Netherlands", *GIS/LIS proceedings 90*, Anaheim, California, Vol 1, pp.58-67
- Mustière, S., Zucker, J., e Saitta, L., (1999), "Cartographic Generalization as a Combination of Representing and Abstracting Knowledge", *Proceedings of the 7th ACM international symposium on Advances in geographic information systems*, ISBN:1-58113-235-2.
- Mustière, S., (2005), "Cartographic generalization of roads in a local adaptive approach: A knowledge acquisition problem", *International Journal of Geographical Information Science*, 19(8-9): 937-955.

- Mustière, S., e Duchêne, C., (2001), “Comparison of different approaches to combine road generalisation algorithms: Galbe, Agent and Cartolearn”, In fourth workshop on progress in automated map generalisation, Beijing, China, 2001. International Cartographic Association, commission on map generalisation and multiple representation, 14p.
- Mustière, S., e Moulin, B., (2002), “What is spatial context in cartographic generalisation?”, Conference Geospatial Theory, Processing and Applications, 2002, IAPRS & SIS, vol. 34, No. 4 Ottawa, Canada. Pages 274-278.
- Nakos, B., Gaffuri, J., e Mustière, S., (2008), “A transition from simplification to generalisation of natural occurring lines”, 11th ICA Workshop on Generalisation and Multiple Representation, 20-21 June 2008, Montpellier, France.
- Nelson, M., e Illingworth, W., (1991), “A Practical Guide to Neural Nets”, Addison-Wesley Publishing Company, Inc., USA, ISBN: 0-201-52376-0/0-201-56309-6, 1991.
- Neun, M., Weibel, R., e Burghardt, D., (2004), “Data Enrichment for Adaptive Generalisation”, In: ICA Workshop on Generalisation and Multiple Representation (Leicester), disponível em <http://ica.ign.fr/>.
- Neun, M., e Burghardt, D., (2005), “Web Services for an Open Generalisation Research Platform”, In: 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña, Spain.
- Nickerson, B., (1988), “Automated Cartographic Generalization for Linear Features”, *Cartographica*, 25 (3), 15-66p.
- Normas de produção cartográfica da série M782, escala 1/50 000 do Instituto Geográfico do Exército, 2001.
- Peitgen, H., et al., (1992), “Chaos and Fractals – New Frontiers of Science”, Institut für Dynamische Systeme, Universität Bremer D-2800 Bremer, Federal Republic of German, 1992.
- Pereira, L., (2011), “Gestão de conhecimento em projectos”, FCA – Editora de informática Lda., ISBN: 978-972722-697-9.

- Peter, B., e Weibel, R., (1999), "Integrating Raster and Vector-based Techniques for the Generalization of Categorical Maps", 19th Int. Cartographic Conference, Ottawa, 74-84.
- Plazanet, C., (1995), "Measurement, Characterization and Classification for automated line feature Generalization", IGN - Service de la recherche - Laboratoire COGIT 2, France, pp 59-68.
- Plazanet, C., (1996), "Enrichissement des bases de données géographiques : Analyse de la Géométrie des objets linéaires pour la generalization cartographique. Application aux routes", Phd Thesis, Universidade de Marne-la-Vallée, 1996. Paris, 185p.
- Plazanet, C., (1997), "Modelling geometry for linear feature generalization", In M. Craglia & H. Couclelis (Eds.), *Geographic Information Research. Bridging the Atlantic* (pp. 264-279). London: Taylor & Francis.
- Plazanet, C., Affholder, J., e Fritsch, E., (1996), "The Importance of Geometric Modelling in Linear Feature Generalisation", IGN - Service de la recherche - Laboratoire COGIT 2, Avenue Pasteur (B.P. 68) 94160 St Mandé Cedex France, 14p.
- Plazanet, C., Bigolin, N., e Ruas, A., (1998), "Experiments with Learning Techniques for Spatial Model Enrichment and Line Generalization", *GeoInformatica*, p.315-333.
- Quinlan, J., (1986), "Induction of Decision Trees", *Machine Learning* 1: 81-106.
- Rabuske, R., (1995), "Inteligência Artificial", Florianópolis (SC), Brasil: Editora da UFSC, 1995. 240p.
- Ratajski, L., (1967), "Phénomène des points de generalization", In: K. Kirschbaum and K. H. Meine (editors), *International Yearbook of Cartography*; Vol.7, pp.143-152. Kirschbaum, Bonn-Bad Godesberg.
- Refenes, N., et al., (1993), "Currency Exchange rate prediction and Neural Network Design Strategies", *Neural Computing & Applications*, Vol 1, no. 1., (1993).
- Regnauld, N., (1998), "Généralisation du bâti: Structure spatiale de type graphe et représentation cartographique", Thèse de doctorat, Université de Provence, Aix-Marseille 1, 1998, 191p.

- Regnauld, N., (2001a), "Constraints based Mechanism to achieve automatic generalization using agent modelling", Proceedings of GIS Research UK 9<sup>th</sup> Annual Conference, pp.329–332, university of Glamorgan, UK.
- Regnauld, N., (2001b), "Contextual Building Typification in Automated Map Generalization", *Algorithmica* (2001) 30: 312–333.
- Reumann, K., e Witkam, A., (1974), "Optimizing curve segmentation in computer graphics", in Proceedings of International Computing Symposium, pp. 467–472.
- Reichenbacher, T., (1995), "Knowledge acquisition in map generalization using interactive systems and machine learning", Proceedings, ACI p 2221-2230 Barcelona, Spain.
- Reichgelt, H., (1991), "Knowledge Representation - an AI Perspective", Norwood, Ablex Publishing Corporation, 1991, 251p.
- Rich, E., (1988), "Inteligência Artificial", São Paulo: McGraw-Hill, 1988.
- Rich, E., e Knight, K. (1991), "Artificial Intelligence", McGraw-Hill, 2<sup>nd</sup> ed., ISBN 0-07-052263-4, 1131p.
- Richardson, D., e Mackaness, W., (1999), "Computacional Processes for Map Generalization", *Cartography and Geographic Information Science*, 26(1), pp. 3-5.
- Rieger, M., e Coulson, M., (1993), "Consensus or Confusion: Cartographers' Knowledge of Generalization", *Cartographica*, 30 (2 & 3), 69-80.
- Rocha, M., Cortez, P., e Neves, J., (2008), "Análise Inteligente de Dados", Tecnologias de Informação, FCA - Editora de Informática, Lda, ISBN 978-972-722-278-0.
- Roberts, J., (1997), "The Development of an Autonomous Self Evaluating Algorithm for Map Generalization: Automated Feature Displacement", Master's thesis, Department of Geography, University of Edinburgh, 1997.
- Robinson, A., (1960), "Elements of Cartography, 2<sup>nd</sup>ed.", New York: Wiley, 343p.
- Ruas, A., (1998), "A Method for building displacement in automated map generalisation", *International Journal on GIS*, 1998. vol12, no 8. 789-803p.

- Ruas, A., (1999), “Modèle de généralisation de données géographiques à base de contraintes et d'autonomie”, PhD thesis, Sciences de l'Information Géographique, Université de Marne La Vallée, 1999, 323p.
- Ruas A., (2000 a), “Project AGENT: Overview and Results of a European R&D Project in Map Generalisation”, ICA Workshop, Barcelona, Spain.
- Ruas, A., (2000 b), “The Role of Meso Objects for Generalization”, Proceedings of the 9<sup>th</sup> Spatial Data Handling Symposium, Beijing, pp. 3b. 50-3b. 63.
- Ruas, A., (2002), “Généralisation et représentation multiple”, Lavoisier Ed., 11 rue Lavoisier 75008 Paris, ISBN 2-7462-0494-0.
- Ruas, A, e Lagrange, J., (1995), “Data and knowledge modelling for generalization”, In Müller J-C, Lagrange J-P, Weibel R (eds), GIS and generalisation: methodological and practical issues. London, Taylor and Francis: 73–90
- Ruas, A., e Plazanet, C., (1996), “Strategies for Automated Generalization”, Proceedings of the 7th Spatial Data Handling Symposium, Delft, the Netherlands, pp. 319-336.
- Rumelhart, D., Hinton, G., e Williams, R., (1986), “Learning internal representations by error propagation”, In D. E. Rumelhart, J. L. McClelland, and 1986. PDP Research Group, Parallel Distributed Processing, Volume 1: Foundations. MIT Press.
- Russel, S., e Norvig, P., (2003), “Artificial Intelligence, a modern approach”, Second Edition, Pearson Education, Inc. Upper Saddle River, New Jersey 07458, ISBN 0-13-080302-2.
- Ruzak, M., e Castner, H., (1990), “Horton's ordering scheme and the generalisation of river networks”, The Cartographic Journal, Vol. 27, No. 2, 1990. 104-112p.
- Sarjakoski, T., e Kilpeläinen, T., (1999), “Holistic Cartographic Generalization by Least Squares Adjustment for Large Data Sets”, In: Proceedings of 19<sup>th</sup> International Cartographic Conference, Ottawa, Canada, pp 1091-1098.
- Sato, Y., (1992), “Piecewise linear approximation of plane curves by perimeter optimization”, Pattern Recognition, 25(12), 1535-1543.

- Sen, A., e Gökgöz, T., (2012), “Clustering approaches for hydrographic generalization”, GIS Ostrava 2012 – Surface models for geosciences, January 23-25, Ostrava.
- Sester, M., (2000), “Generalization Based On Least Squares Adjustment”, International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B4, pp. 931-938, Amsterdam 2000.
- Shea, K., e McMaster, R., (1989), “Cartographic Generalization in a Digital Environment: When and How to Generalize”, In: Proceedings for Auto-Carto 9, Baltimore, MD, 1989:56-67.
- Shi, W., e Cheung, C., (2006), “Performance evaluation of line simplification algorithms for vector generalization”, The Cartographic Journal, Vol. 43, No 1 pp. 27-44.
- Silva, A., e Videira, C., (2001), “UML metodologias e ferramentas CASE”, Centro Atlântico, Lda., 2001, Portugal, ISBN: 972-8426-36-4.
- Simon, H., (1983), “Why should machines learn?”, In Machine Learning, An Artificial Intelligence Approach, volume 1, 28–3. Springer-Verlag, 1983.
- Skopeliti, A., e Tsoulos, L., (1999), “On the parametric description of cartographic line shape”, Cartographica 36 (3), pp. 57-69.
- Smith B., Jackson H., e Maier L., (1999), “The Future of Automated Map Generalization”, Intergraph Corporation, Huntsville, Alabama, USA, 8p.
- Steiniger, S., e Meier, S., (2004), “Snakes: a technique for line smoothing and displacement in map generalisation”, In 7<sup>th</sup> ICA workshop on generalisation and multiple representation, Leicester, GB, 11p.
- Steiniger, S., e Weibel, R., (2005), “A Conceptual Framework for Automated Generalization and its Application to Geologic and Soil Maps”, Proceedings 22nd International Cartographic Conference, La Coruña (Spain), 11-16 July 2005. CD-ROM.
- Steiniger, S., e Weibel, R., (2007), “Relations among map Objects in Cartographic Generalization”, Cartography and Geographic Information Science, vol 34, N°3, pages 175-197.



- Steiniger, S., e Taillandier, P., e Weibel, R., (2010), “Utilising urban context recognition and machine learning to improve the generalisation of buildings”, *International Journal of Geographical Information Science*, Vol. 24, No. 2, February 2010, 253–282.
- Surkan, A., e Singleton, J., (1990), “Neural networks for bond rating improved by multiple hidden layers”, *Neural Networks, 1990 IJCNN International Joint Conference on*, 157-162.
- Szego, J., (2011), “Classifying, Analysing and Experiencing maps – A tentative humanistic approach”, *Lecture notes in Geoinformation and Cartography, Advances in Cartography and GIScience Volume1* 33-51. Springer, 2011.
- Taillandier, P., Duchêne, C., e Dogoul, A., (2009), “Using Belief Theory to Diagnose Control Knowledge Quality, Application to cartographic generalization”, *Computing and communication Technologies, 2009. RIVF '09. International Conference on Computing and Telecommunication Technologies. 13-17 July 2009*, ISBN: 978-1-4244-4566-0.
- Taillandier, P., e Taillandier, F., (2012), “Multi-criteria diagnosis of control knowledge for cartographic generalization”. *European Journal of Operational Research*, Elsevier, 2012, 217 (3), pp.633-642.
- Thompson, R., e Richardson, D., (1995), “A Graph Theory Approach to Road Network Generalization”, In: *Proceedings of the 17<sup>th</sup> ICA/ACI Conference, Barcelona, 1995*. 1871-1880p.
- Töpfer, F., e Pillewizer, W., (1966), “The Principles of Selection”, *The Cartographic Journal*, Vol. 3, No. 1, pp. 10-16.
- Vázquez, J., (2003), “Desarrollo de estrategias, métodos y herramientas para la automatización y mejora de los procesos de producción cartográfica. Aplicación a la generación de mapas excursionistas.”, *Phd Thesis, Universidad Politécnica de Valencia, Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría, Valencia, Spain*, 458p.
- Vianna, C., (1997), “Generalização Cartográfica em Ambiente Digital Escala 1:250000 a partir de Dados Cartográficos Digitais na Escala 1:50000”, *Dissertação*

- (Mestrado em Engenharia Cartográfica) - Departamento de Engenharia Cartográfica, Instituto Militar de Engenharia, Rio de Janeiro, 1997, 242p.
- Visvalingham, M., e Williamson, P., (1995), "Simplification and Generalisation of Large-Scale Data for Roads: a comparison of two filtering algorithms", *Cartography and Geographical Information Systems* 22 (4), 264 - 275.
- Visvalingham, M., e Whyatt, J., (1993), "Line Generalisation by Repeated Elimination of Points", *Cartographic Journal.*, 30 (1), 46 – 51.
- Wang, Z., e Muller, J., (1998), "Line generalisation based on an analysis of shape characteristics", *Cartography and Geographic Information Systems*, 25(1), 3-15.
- Wasserman, P., (1989), "Neural Computing, Theory and Practice", Van Nostrand Reinhold, New York, p 230.
- Weber, R. (1998) "Intelligent Jurisprudence Research", Doctoral Dissertation, Graduate Program of Production, Engineering at the Federal University of Santa Catarina, Brazil. May, 1998, 161p.
- Weber, E., Lisboa, J., Iochpe, C., e Hrsenack, H., (1998), "Geospatial metadata in Brazil: an experience in data documentation of an environmental GIS application", In: INT. Conference & Exhibition on Geographic Information - GISPlanet, 1998, Lisbon, Portugal. Lisbon: USIG, 1998.
- Weibel, R., (1991), "Amplified Intelligence and Knowledge-Based Systems", In: Buttenfield, B.P. and McMaster, R.B. (eds.). *Map Generalization: Making Rules for Knowledge Representation*. London: Longman, 172-186.
- Weibel, R., (1995), "Three essential building blocks for automated generalization", Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland.
- Weibel, R., (1997), "Generalization of Spatial Data – Principles and Selected Algorithms", In: Van Kreveld, M., Nievergelt, J., Roos, Th. and Widmayer, P. (eds.): *Algorithmic Foundations of GIS. Lecture Notes in Computer Science*, Berlin: Springer-Verlag, Vol. 1340: 99-152.
- Weibel, R., e Dutton, G., (1999), "Generalizing Spatial Data and Dealing with Multiple Representations", In: Longley, P., Goodchild, M.F., Maguire, D. J. and Rhind,

- D.W. (eds.): Geographical Information Systems: Principles, Techniques, Mangement and Application, 2nd Edition. Chichester: Wiley, p. 125-155.
- Weibel, R., Keller, S., e Reichenbacher, T., (1995), “Overcoming the Knowledge Acquisition Bottleneck in Map Generalization : the Role of Interactive Systems and Computationnal Intelligence”, Proceedings of COSIT’95, p. 139-156.
- Woojin, P., e Kiyun, Y., (2011), “Hybrid line simplification for cartographic generalization”, Pattern Recognition Letters 32 (2011) 1267–1273.
- Wooldridge, M., (2002), “An Introduction to MultiAgent Systems”, Department of Computer Science, University of Liverpool, Uk, ISBN 0-471-49691-X.
- Yohannes, Y., e Hoddinott, J., (1999), “Classification and Regression Trees: An Introduction”, International Food Policy Research Institute, 2033 K Street, N. W. Washington, D.C. 20006 U.S.A.
- Zhang, L., e Tian, Z., (1997), “Refinement of Douglas-Peucker Algorithm to move the segments towards only one side”, Proceedings of 18<sup>h</sup> International Cartographic Conference, Stockholm, Vol 2, Sweden, pp 830-835.